GfK Verein

WORKING PAPER SERIES

HOW VALID ARE RESPONSE-TIME MEASURES FOR CAPTURING IMPLICIT BRAND ATTITUDES?

Anja Dieckmann, Matthias Unfried, Regina Schreder, Kathrin Kissel

WORKING PAPER /// NO. 7 / 2018

How valid are response-time measures for capturing implicit brand attitudes?

Anja Dieckmann*† Matthias Unfried† Regina Schreder‡
Kathrin Kissel§

Abstract— *In an online questionnaire study on three different product categories–insurances, shampoo, hair coloration–we compare two methodological approaches for measuring brand attitudes: (a) Standard rating scales: Respondents indicate on a 7-point scale how well a certain attribute fits a brand, and (b) simple response-time measure: Respondents indicate if a certain attribute fits a brand, yes or no. Speed of response is taken as a measure of fit between attribute and brand. Both methods are compared on several criteria: (1) Ability to discriminate between different attributes and brands, (2) evaluation of comprehensibility and fun of use by respondents, (3) agreement with the established Implicit Association Test (IAT; Greenwald et al., 1998), (4) validity in terms of correlation with marketing KPIs, and (5) test-retest reliability. While the simple response-time measure produces higher variance between brands and items, rating scales show higher retest reliability, and higher correlation with stated purchase and recommendation intentions as well as purchase likelihood inferred from a choice-based procedure. Moreover, the simple response-time measure does not show higher agreement with the IAT compared to rating scales. Thus, the alleged higher differentiation that is suggested by the response-time measure is attributed to greater error variance rather than reliable attitude differences. Merely in terms of fun of use did the simple response-time measures perform better than rating scales.*

Keywords— *Emotion Capturing; Facial Coding; Affective Computing; Ad Testing; Intercultural Studies; China*

1 Introduction

With the advent of neuroscientific approaches in marketing and consumer research, implicit processes—processes consumers may not be aware of and therefore cannot deliberately report—have received increased attention (e.g., Bridger, 2015; Gattol et al., 2011; Scheier and Held, 2006). As neuroscientific methods such as EEG and fMRI are very cost- and time intensive, and even heart rate and skin conductance measurements cannot be applied in online questionnaires due to their dependence on sensor equipment, implicit measures based on response times may offer an efficient alternative route to capture implicit attitudes (Scheier, 2006). These indirect question techniques augur access to consumers' unconscious attitudes, but without the need for invasive physiological measurements and expensive equipment.

Interest in such implicit measurements was further boosted by the tremendous success of Kahneman's proposed dual-processing theory in the wake of him winning the Nobel Memorial Prize in Economic Sciences in 2002, summarized in the best-selling book "Thinking, fast and slow" (Kahneman, 2011). His proposed fast, intuitive, emotional and often unconscious "System 1" has received much attention also by marketing practitioners (summarized in The Economist, 2013). It renewed interest in dual-process theories, which have been proposed decades ago to account for people's sometimes seemingly

*Corresponding author, anja.dieckmann@gfk-verein.org

†GfK Verein, Fundamental Research, Nuremberg, Germany

‡University of Bayreuth, Chair of Marketing and Consumer Behavior, Germany

§GfK SE, Nuremberg, Germany

Citation: Anja Dieckmann, Matthias Unfried, Regina Schreder, Kathrin Kissel (2018). How valid are response-time measures for capturing implicit brand attitudes?. GfK Working Paper Series, No. 7 / 2018.

irrational decisions or for discrepancies between stated attitudes and behavior (e.g., Petty and Cacioppo, 1986; Strack and Deutsch, 2004). As a result, there is great demand in consumer research for tools to capture such implicit decision-making processes and attitudes that consumers may be unaware of, making them difficult to access with traditional survey approaches. Implicit measurement techniques may help to fill this blank.

Implicit attitude measurement

Several implicit measurement approaches have been proposed in psychology for capturing attitudes. What they have in common is that the purpose of the question is not evident to the respondent. "[They] focus a participant's attention on performing some task that can indirectly reveal the underlying construct such that inaccessible and closely held attitudes can be measured" (Brunel et al., 2004, p. 387). To interpret the outcomes of such implicit methods, it is therefore necessary to make inferences about attitudes, often based on differences in response times. The underlying assumption is that attitudes are stored in associative networks (Fazio, 2001; Greenwald et al., 2002). Perception of an attitude object (e.g., a brand logo) will activate associated content. The stronger the association, the faster this content will come to mind, hence the focus on response times. It should be noted that implicit measurement targets both, attitudes one is unaware of and attitudes one is aware of but does not want to openly admit (e.g., prejudices; Brunel et al., 2004). Thus, implicit measurement is often used to study prejudices and negative stereotypes about, e.g., people with certain characteristics such as nationality, gender, sexual orientation, skin color, age, body weight – attitudes respondents are reluctant to honestly report or even admit to themselves for social desirability reasons (for several examples, see https://implicit.harvard.edu/implicit/). The motivation for applying implicit methods is that—because question purpose is hidden—respondents cannot bias their answer towards greater social desirability.

One of the most prominent instruments for implicit measurement is the Implicit Association Test (IAT Greenwald et al., 1998). In a series of choice tasks, the IAT measures the relative difference in association of two target concepts (such as brands, for instance) with an evaluative concept (e.g., pleasantness). The evaluation concept is presented as a set of, for example, pleasant and unpleasant words, to be categorized in a two-choice task by pressing designated response keys on a keyboard. The two target concepts are also to be categorized in a dual task. If, for instance, one target concept is positively valued and the other is negatively valued, systematic response time differences can be observed depending on which key is to be used for categorization. When the positively valued target concept is to be categorized with the key that is also used for categorizing pleasant words, response times tend to be faster than when having to use the key for unpleasant words. The opposite holds for the negatively valued target concept.

There are two popular versions of the IAT. The most familiar, original version is the evaluative IAT (Greenwald et al., 1998) that uses categorization of pleasant vs. unpleasant words along with categorization of target concepts. It is the version frequently used to study prejudices (e.g., Cunningham et al., 2003, 2004; Gawronski and Conrey, 2004; McConnell and Leibold, 2001) and indicates which of two attitude objects is valued more positively or is preferred. Another version is the self vs. other IAT (Greenwald et al., 2002) used to study which of two constructs is associated more closely with one's concept of self. This version of the IAT has already been used for capturing consumer-brand relationships (Brunel et al., 2004).

For a latency-based psychometric measure, the IAT has demonstrated relatively high reliability and validity (Bosson et al., 2000; Hofmann et al., 2005; Nosek et al., 2007; Lane et al., 2007). However, the effort required to conduct an IAT to provide as output a merely relative evaluation of two concepts on one dimension is high (time to complete is about 5 to 10 min, Greenwald et al., 2003). This poses an important disadvantage for practitioners in consumer research.

Direct attitude measurement using response times

Given the limitations of the IAT and other established implicit attitude tests from academic psychology, it comes as no surprise that practitioners have started to propose shortcuts. One of the most straight-forward approaches is to ask whether a certain attribute fits a target concept such as a brand, yes or no (e.g., Heinsen and Lorenz, 2011; Scheier, 2006; Schmidt et al., 2015).

Time to respond is treated as a measure of associative strength, with fast yes-responses indicating a closer association between brand and attribute. While acknowledging that response-time measures taken while respondents answer direct yes-no questions are not truly implicit, the underlying expectation seems to be that they at least foster spontaneous answers. Because of this spontaneity, it may be possible to capture unconscious attitudes to some extent and as a result get closer to results from full-blown implicit tests than traditional rating scales. We will test this assumption using two different item sets for capturing brand attitudes.

Brand associations and brand attachment

Questionnaires assessing different aspects of brands are ubiquitous in consumer research and there exist many different item batteries, from brand image (e.g., Cho et al., 2015) over brand commitment (e.g., Shuv-Ami, 2012) to brand personality scales (e.g., Aaker, 1997), just to mention a few. We focus on two concepts for which implicit testing has been proposed or seems particularly suitable: need-based brand associations and brand attachment.

Based on the assumption that brands satisfy basic needs in consumers, Scheier and Held (2007, 2010) draw on influential motivation theories in psychology (espe-

cially The Zurich Model of Social Motivation by Bischof, 1993) and postulate a system of six basic needs: Discipline, Safety, Pleasure, Stimulation, Adventure, Autonomy. Via so-called "codes" (that can be words, but also colors, shapes, sounds, etc.) in their marketing communication, marketers try to associate their brands with those needs. Scheier (2006, 2008) himself is a proponent of implicit approaches for capturing brand image and the needs a brand addresses using verbal item lists and picture sets. In our study, we use an item list addressing these need-based brand associations.

However, one may argue that consumers may often be aware of a brand's associated needs, as they are often stated in brand communication in terms of the benefits a brand provides. Moreover, consumers have little reason to not openly report on the associations when explicitly asked about them. In consequence, explicit and implicit measure of the associative strength may often show little differences.

In contrast, consumers may often not be fully conscious of more emotional facets of consumer-brand relationships. Also, people may be reluctant or feel awkward to explicitly talk about these. Fournier (1998), for instance, has intensively studied the relationships consumers have with brands in in-depth and in-home qualitative interviews, acknowledging that aspects that are central to one's core concept of self may act below the level of conscious awareness. Thus, consumer-brand relationships seem to particularly lend themselves to implicit testing. To capture the strength of such relationships, the concept of brand attachment has been invoked (Fournier, 2009; Thomson et al., 2005). Based on work on interpersonal attachment, Thomson et al. (2005), in line with Fournier (1998), distinguish the concept of brand attachment from attitudes by the often highly affective link between brand and the self. Thomson et al. (2005) have developed a scale for capturing brand attachment consisting of emotional items that represent three core dimensions of the construct: Affection, Passion, Connection. We use these items in our study.

Research objective

We compare traditional rating scales for capturing attitudes about brands with response-time measures derived from yes-no questions. To this end, we ask participants to answer questions on brand attachment or brand associations either by a traditional 7-point rating scale or by dichotomous yes-no answers with response time as measurement. The two methods are compared using the following criteria: (1) Ability to discriminate between different attributes and brands, (2) evaluation of comprehensibility and fun of use by respondents, (3) agreement with the established Implicit Association Test (IAT Greenwald et al., 1998), (4) validity in terms of correlation with marketing KPIs, and (5) test-retest reliability.

2 Method

Participants

Participants were recruited from the German GfK Online Panel via email. Panelists were screened for product category relevance and brand knowledge. Overall, 1425 eligible participants completed the questionnaire, 58.7% female, with an average age of 42.0 years. One week after the first questionnaire, invitations to participate in a retest of the main parts of the questionnaire were sent out to the participants of the first questionnaire. A subset of 1201 of the participants completed the retest (58.3% female, 42.2 years).

Procedure

Initially, panelists interested in participating in the questionnaire study were screened for relevance of at least one of the following three different target product categories: insurances, hair colorations, shampoos. Also, they had to know at least two brands from our respective brand lists that consisted of the top-eight brands in terms of market share in Germany. In particular, panelists were only eligible to participate if they indicated: (a) to either own or plan to purchase a private health insurance and to have insurance contracts with at least two brands from our insurance brand list, or (b) to buy hair coloration at least once every 4 to 6 months and to know at least two brands from our hair coloration brand list, or (c) to buy shampoo at least once every 4 to 6 months and to know at least two brands from our shampoo brand list.

Participants were then assigned to one product category for which they fulfilled the screening criteria. Participants had to evaluate two brands that they had indicated to know in the initial screening questions. How they indicated their evaluations varied along two factors: (1) question format: Item ratings or dichotomous yes-no assignment of items with response-time measurement; (2) evaluation dimensions: brand associations (16 items) or brand attachment (14 items). Participants in the brand associations conditions also completed the evaluative IAT with pleasant vs. unpleasant words, and participants in the brand attachment conditions completed the self-other IAT. The two-factorial design resulted in four conditions, to which participants were randomly assigned (see Figure 1).

The questionnaire started with four short questions about involvement in the assigned product category to focus participants' attention. Then followed the questions on brand associations and attachment.

To measure consumers' associations with a brand, each of the six needs proposed by Scheier and Held (2010) was operationalized by two items: (1) discipline: *dutiful, disciplined;* (2) safety: *approved, reliable;* (3) pleasure: *easygoing, delightful;* (4) stimulation: *lively, creative;* (5) adventure: *thrilling, adventurous;* (6) autonomy: *respected, powerful.* To complement the list, we added four items used in the short version of the brand potential index BPI, formerly known as Brand Assess-

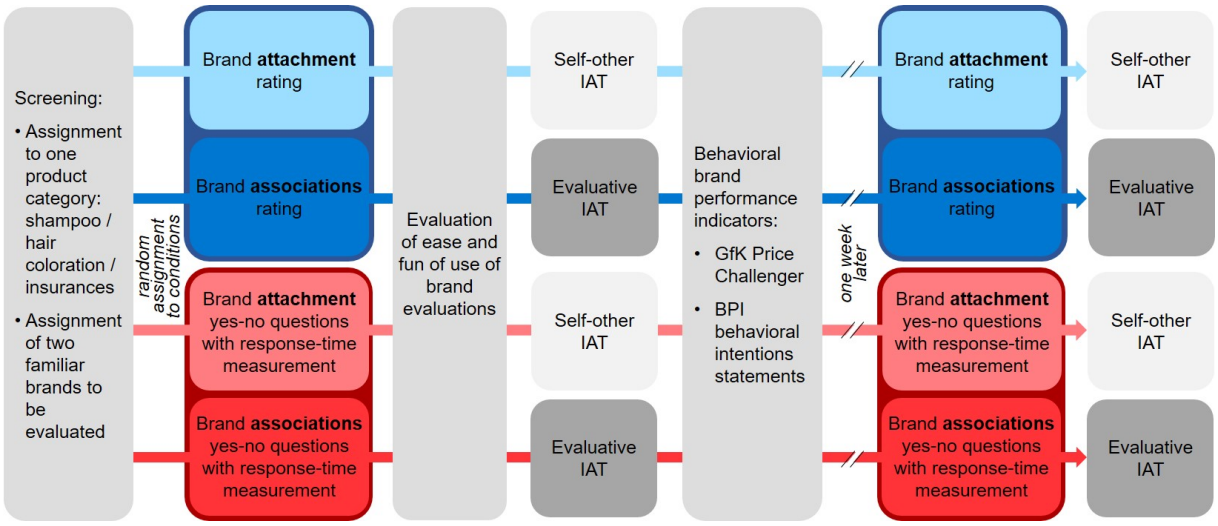

Figure 1: Schematic of interview flow and design. To highlight our main factor of interest, the question format, the two rating conditions are marked in blue, and the two yes-no conditions with response-time measurement in red.

ment System (BASS), by GfK (Högl and Hupp, 2004): *superior, unique, valuable, likable.*[1] These 16 items represented our brand associations item set. All items were positively valued, so we expected brands that score high on these items to achieve more positive scores in the evaluative IAT relative to other brands.

For brand attachment, we used the item list introduced by Thomson et al. (2005) to capture three core emotional dimensions: (1) affection: *affectionate, friendly, loved, peaceful;* (2) passion: *passionate, delighted, captivated* (3) connection: *connected, bonded, attached.*[2] In addition, work from our own research group highlights the outstanding importance of trust in consumer-brand relationships (e.g., Gaspar, 2011). To account for that, we added the following four trust items: *honest, dependable, trusting, fair.*[3] These 14 items represented our brand attachment item set. We expected that brands that score high on these items to be more closely associated with self-related words in the self-other IAT than other brands.

The participants evaluated the two assigned brands in the format and with the items of their respective conditions. The two brands were alternated, and the items randomized separately for each brand. The question text and one of the two assigned brands' logos always appeared on the top of the screen, and one second later, the to-be-answered item appeared below the logo together with the answer scale. The questions were phrased as follows in the different conditions:

For brand associations ratings, participants were asked: How strongly do you associate the following term with [brand logo]? Then an item from the brand associations list (*dutiful, disciplined, approved, reliable, easy-going, delightful, lively, creative, thrilling, adventurous, respected, powerful, superior, unique, valuable, likable*) appeared together with a 7-point answers scale from 1 (labeled "not at all") to 7 ("very strongly"). Points between poles were not labeled. Answers were given by mouse click.

For brand associations dichotomous yes-no questions with response-time measurement, participants were asked: Do you associate the following term with [brand logo]? Then an item from the brand associations list appeared together with the answer alternatives "yes" and "no". Answers were given by key press of "A" for "yes" and "L" for "no". A note indicating the assignment of keys to the answer alternatives was always visible at the bottom of the screen.

For brand attachment ratings, participants were asked: How well does the following term describe your relationship with [brand logo]? Then one of the brand attachment items (*affectionate, friendly, loved, peaceful, passionate, delighted, captivated, connected, bonded, attached, honest, dependable, trusting, fair*) appeared together with a 7-point answers scale from 1 (labeled "not at all") to 7 ("very well"). Points between poles were not labeled. Answers were given by mouse click.

For brand attachment dichotomous yes-no questions with response-time measurement, participants were asked: Does the following term describe your relationship with [brand logo]? Then one of the brand attachment items appeared together with the answer alternatives "yes" and "no". Answers were given by key press of "A" for "yes" and "L" for "no". A note indicating the assignment of keys to the answer alternatives was always visible at the bottom of the screen.

Before evaluating brands, participants had one round of practice in which they evaluated their hometown and their last holiday destination on four positively values attributes not included in the test item sets (*beautiful, exciting, curious, harmonious*).

[1] German questionnaire items used in the study are available on request.

[2] German translation of the items is available on request.

[3] German questionnaire items used in the study are available on request.

Immediately after the brand evaluations, participants evaluated ease and fun of use of the way in which they just had answered the questions on brands. Then they completed an IAT. Participants in the brand associations condition completed the evaluative IAT with positively (*happy, peace, love, joy, wonderful*) or negatively (*agony, evil, awful, horrible, nasty*) valued words (selected from Nosek et al., 2002, and translated into German)[4], and logos as well as product pictures of the two assigned brands. Participants in the brand attachment condition completed the self-other IAT with self (*I, self, my, me, own*) or other (*they, them, your, you, other*) words (Asendorpf and Mücke, 2002)[5], and logos as well as product pictures of the two assigned brands. The standard seven-block version of the IAT was applied (cf. Greenwald et al., 2003, Table 1), with 40 trials per test block (i.e., blocks 4 and 7), and 20 for all other blocks.

After the IAT, participants completed the GfK Price Challenger (GPC Wildner, 2003), a choice-based procedure in which participants choose repeatedly from a set of branded products whose prices vary between choice tasks. All brands that a participant had indicated to know in the initial screen questions were included in the GPC. Participants were exposed to eleven choice tasks and had to select the product they would purchase given the presented price constellations in each task. Via a logistic model, individual purchase probabilities per brand can be estimated from the observed choices.

Subsequently, participants answered questions about brand-related behavioral intentions for all brands that they had indicated to know. Items correspond to the behavioral statements of the brand potential index BPI (Högl and Hupp, 2004): (a) *I use the brand* [name] *whenever I have the opportunity* (only for product categories hair coloration and shampoo); (b) [name] *is a brand for which I am willing to spend more than for other brands;* (c) [name] *is a brand I am more likely to recommend than other brands;* (d) [name] *is a brand I am more likely to buy in the future than other brands.*

The questionnaire ended with some visual analogue scales for brand evaluation (not reported here) and sociodemographic questions.

One week after the first questionnaire had been distributed, participants were invited to the retest. For the retest, participants were assigned to the same product categories, brands, and conditions. Only the brand evaluations, in the respective question formats and with the respective items, as well as the IAT were re-administered in the retest.

[4]German translation of the items is available on request.

[5]The authors kindly provided their original German items. They are available on request.

3 Data preparation of response times in dichotomous yes-no questions

Response times are particularly noisy and have to be treated with a certain level of caution. Different options for dealing with the relatively low signal-to-noise ratio have been discussed in the literature (e.g., Fazio, 1990; Mayerl and Urban, 2008), including elimination of data with extreme response times. However, Fazio (1990) argues for caution when it comes to such eliminations and that one needs to keep in mind that the distribution of response times is inevitably skewed. How to deal with outliers depends on how response times are used. Since we intend to interpret response times as a measure of associative strength, we mainly refer to standards proposed for the IAT and its variants (e.g., the Brief IAT; Nosek et al., 2014). In particular, it needs to be decided how to treat extremely short response times, which could indicate a lack of processing of the question content, and extremely long response times, which hint to a failure to answer spontaneously as instructed. Most authors have focused on truncation/recoding to boundary values, elimination of single responses, and elimination of respondents who exhibit too short or too long response times on a certain proportion of items (Glashouwer et al., 2013; Greenwald et al., 2003; Nosek et al., 2014). Based on multiple evaluation criteria such as sensitivity, internal consistency and agreement with other measures of the same topic, Nosek et al. (2014) report only small differences between different thresholds; generally, evaluation criteria tended to slightly improve when extreme response times were recoded to boundary values rather than eliminated, and when respondents with extreme response times in a significant subset of trials were excluded rather than included.

In line with these recommendations, we removed respondents from the dataset whose response times are below 300 ms in 20% or more of the items (thus being a bit more conservative than general IAT rules, according to which respondents with 10% or more response times shorter than 300 ms are removed; Greenwald et al., 2003). We interpret such a response pattern as an attempt to click through the questionnaire as quickly as possible without reading question contents. Moreover, we remove respondents with 50% or more response times longer than 3500 ms because this means that less than half of the items have been answered spontaneously and no meaningful brand comparison can be conducted. Application of these criteria led to the following exclusions: We excluded 22 respondents (1.5%) whose response times were too short (below 300 ms in 20% of the items) and 29 respondents (2.0%) whose response times were too long ($>$ 3500 ms for 50% or more items).

For the remaining respondents, we recode response times below 300 to 300 ms, and response times above 3500 to 3500 ms. The upper limit was chosen because in our study respondents received a reminder to answer spontaneously at 3500 ms. Note that recoding to an up-

per boundary of 2000 ms produced the best outcomes in Nosek et al. (2014), but differences compared to a boundary of 4000 ms were very small. Overall, 139 items (out of 19804 items, i.e., 0.7%) were recoded to 300 ms, and 41 items (i.e., 0.2%) were recoded to 3500 ms.

Moreover, to compare the response times in dichotomous yes-no questions with the degree of agreement indicated by rating scale answers, it is necessary to distinguish between yes and no answers. Based on the assumption of response times signaling associative strength, fast yes answers should be indicative of strong agreement that brand and item are associated, whereas slow response times should indicate weak agreement. Likewise, for no answers, slow responses should indicate weak disagreement, and fast responses strong disagreement. To derive one continuous measure of agreement from response times, we propose the following transformations:

$$RT_{trans} = \begin{cases} -1 \cdot \left(\frac{RT}{1000} - 3.5\right) & \text{for "yes" answers} \\ \frac{RT}{1000} - 3.5 & \text{for "no" answers} \end{cases}$$

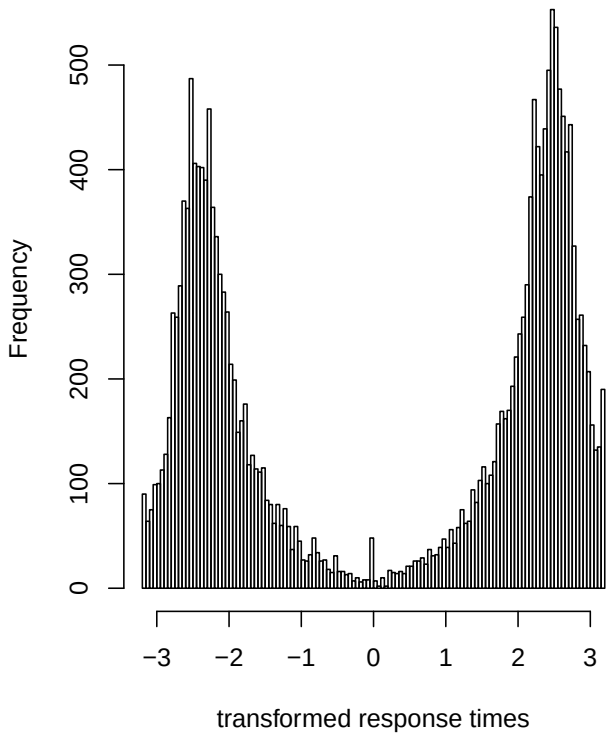

Figure 2: Distribution of transformed response times for dichotomous yes-no answers for both attachment and associations items

As response times below 300 ms have been recoded to 300 ms, transformed response times range from -3.2 to 3.2. Figure 2 shows the distribution of the transformed response times. Very slow yes and no responses end up close to zero, reflecting undecidedness in terms whether or not brand and item are associated. This transformed response-time measure (RT measure) can now be compared to the ratings.

In addition, we treat the top-4 boxes on our 7-point rating scales as agreement[6] and can thus compute percentage of agreement per brand and item. For the dichtomous yes-no condition, we do the same by ignoring response times and simple computing percentage of yes answers per brand and item. Thus, we also compare ratings and yes-no answers in terms of percentage top-4 boxes and percentage yes answers.

We will evaluate dichotomous yes-no questions with response-time measurement and ratings on five criteria:

- Differentiation between attributes and brands
- Evaluation of ease and fun of use by respondents
- Validity: Agreement with IAT results
- Validity: Agreement with GPC results and behavioral statements
- Test-retest reliability

Apart from test-retest reliability, all criteria are evaluated based on the first measurement point.

4 Results

Ability to discriminate between different attributes and brands

Rating scales often lead to flat attribute profiles for brands, that is, respondents seem to evaluate a given brand quite similarly on a range of different items (e.g., a brand is rated quite high or quite low on all dimensions, which could be due to halo effects; Thorndike, 1920). Likewise, there is often little differentiation between brands on a given item (e.g., all brands are rated as quite good in terms of trustworthiness; cf. the positivity bias often found in customer satisfaction surveys, Peterson and Wilson, 1992).

Due to the more spontaneous nature of dichotomous yes-no answers with response-time measurement, we expect more differentiation in that format, revealing differences between brand characteristics as well as between brands that are only expressed spontaneously, but dismissed when rating scales require a numerical judgment, probably leading to rationalization of one's answers or unwillingness to explicitly indicate strong differences.

a) Variability between different attributes of a brand

First, we explore whether our RT measure leads to a more differentiated attribute profile for the brands. Per brand, we compute the mean rating for each item and explore how much the mean ratings vary across all items.

[6] There are two reasons for treating top-4 boxes as agreement: First, the answer scale was unipolar, starting at 1=not at all, so that even the middle can be interpreted as partial agreement, and second, percentages of top-4 boxes were comparable in size to percentages of yes-answers in the dichotomous conditions.

	Variability of evaluations of a brand across items resulting from different measurement approaches:				
	Variance of the mean ratings across all attachment items		Variance of the mean RT measure across all attachment items	Variance of the mean ratings across all associations items	Variance of the mean RT measure across all associations items
MEAN VARIANCE	*0.32*	<	***0.78******	*0.48*	< ***1.04*****
Insurance A	0.33		**0.82**	0.75	**1.39**
Insurance B	0.43		**1.38**	0.67	**1.56**
Insurance C	0.26		**1.71**	1.12	**1.70**
Insurance D	0.72		**1.62**	1.23	**1.61**
Insurance E	0.34		**1.36**	0.63	**0.91**
Insurance F	0.65		**0.85**	1.07	**1.30**
Insurance G	0.31		**0.75**	**0.66**	0.64
Insurance H	0.56		**0.90**	0.80	**2.21**
Hair Coloration A	0.32		**0.61**	0.27	**0.85**
Hair Coloration B	0.13		**0.40**	0.17	**0.75**
Hair Coloration C	0.13		**0.59**	0.35	**0.93**
Hair Coloration D	0.26		**0.36**	0.24	**0.76**
Hair Coloration E	0.11		**0.63**	0.30	**0.96**
Hair Coloration F	0.25		**0.57**	0.06	**0.82**
Hair Coloration G	0.24		**0.48**	0.38	**1.09**
Hair Coloration H	0.21		**0.49**	0.09	**0.88**
Shampoo A	0.61		**1.19**	0.48	**0.68**
Shampoo B	0.26		**0.47**	0.30	**0.77**
Shampoo C	0.18		**0.51**	0.31	**0.47**
Shampoo D	0.28		**0.58**	0.36	**1.33**
Shampoo E	0.15		**0.82**	0.41	**1.27**
Shampoo F	0.19		**0.36**	0.33	**0.69**
Shampoo G	0.22		**0.86**	0.50	**0.93**
Shampoo H	**0.51**		0.31	0.16	**0.47**

Table 1: Variances per brand across mean attachment and associations items. For easier comparison between ratings and RT measure, higher variances are highlighted in bold. Asterisks mark a significant difference between mean variances: *** p<.001.

	Variability of agreement proportions for a brand across items resulting from different measurement approaches:				
	Variance of proportion of top-4 boxes across attachment items		Variance of proportion of yes answers across attachment items	Variance of proportion of top-4 boxes across associations items	Variance of proportion of yes answers across associations items
MEAN VARIANCE	*.022*	<	***.039******	*.025*	< ***.052*****
Insurance A	.031		**.041**	.034	**.067**
Insurance B	.032		**.064**	.050	**.082**
Insurance C	.033		**.094**	.050	**.085**
Insurance D	.038		**.075**	.054	**.075**
Insurance E	.020		**.063**	.035	**.042**
Insurance F	.037		**.040**	.045	**.060**
Insurance G	.018		**.028**	**.039**	.030
Insurance H	.026		**.048**	.050	**.107**
Hair Coloration A	.015		**.027**	.015	**.042**
Hair Coloration B	.010		**.024**	.012	**.038**
Hair Coloration C	.012		**.029**	.014	**.046**
Hair Coloration D	**.020**		.017	.010	**.035**
Hair Coloration E	.004		**.028**	.012	**.048**
Hair Coloration F	.019		**.031**	.003	**.047**
Hair Coloration G	.025		**.026**	.021	**.051**
Hair Coloration H	**.027**		.025	.003	**.042**
Shampoo A	.043		**.058**	.030	**.042**
Shampoo B	.014		**.016**	.018	**.039**
Shampoo C	.015		**.031**	.017	**.030**
Shampoo D	.014		**.032**	.017	**.064**
Shampoo E	.008		**.045**	.013	**.064**
Shampoo F	.013		**.019**	.018	**.032**
Shampoo G	.014		**.048**	.023	**.046**
Shampoo H	**.032**		.018	.014	**.025**

Table 2: Variances of proportion of top-4 respectively yes answers per brand across attachment and associations items. For easier comparison of top-4 and yes proportions, higher variances are highlighted in bold. Asterisks mark a significant difference between mean variances: *** p<.001.

	Variability of item scores across brands resulting from different measurement approaches:								
	INSURANCE			HAIR COLORATION			SHAMPOO		
	Variance of mean rating across brands		Variance of mean RT measure across brands	Variance of mean rating across brands		Variance of mean RT measure across brands	Variance of mean rating across brands		Variance of mean RT measure across brands
MEAN VARIANCE (attachment)	*0.15*	<	***0.27*****	*0.11*		***0.16***	0.13	<	***0.28*****
affectionate	0.18		**0.25**	0.06		**0.33**	0.14		**0.31**
friendly	**0.19**		0.16	0.05		**0.12**	0.09		**0.35**
loved	0.12		**0.49**	0.07		**0.20**	0.15		**0.27**
peaceful	**0.17**		0.13	**0.12**		0.06	0.18		**0.50**
passionate	0.18		**0.22**	0.17		**0.37**	0.12		**0.39**
delighted	0.13		**0.20**	0.08		**0.19**	0.09		**0.14**
captivated	0.18		**0.21**	**0.14**		0.11	**0.19**		0.08
connected	0.07		**0.42**	**0.22**		0.20	**0.10**		0.05
bonded	0.17		**0.40**	**0.09**		0.07	**0.13**		0.10
attached	0.13		**0.21**	**0.10**		0.06	**0.17**		0.16
honest	0.13		**0.23**	0.05		**0.07**	0.08		**0.35**
dependable	0.11		**0.24**	**0.21**		0.11	0.10		**0.34**
trusting	0.12		**0.18**	0.11		**0.17**	0.15		**0.27**
fair	0.15		**0.45**	0.10		**0.11**	*0.18*		**0.62**
MEAN VARIANCE (association)	*0.17*	<	***0.34****	*0.10*	<	***0.17*****	*0.18*	<	***0.32*****
dutiful	0.07		**0.68**	0.11		**0.38**	0.06		**0.20**
disciplined	**0.18**		0.09	0.11		0.11	**0.12**		0.10
approved	0.10		**0.16**	**0.10**		0.08	0.40		**0.47**
reliable	0.12		**0.40**	**0.11**		0.07	**0.16**		0.14
easygoing	**0.15**		0.11	0.09		**0.18**	0.08		**0.28**
delightful	0.15		**0.45**	0.05		**0.08**	**0.23**		0.20
lively	**0.25**		0.24	0.11		**0.31**	0.18		**0.37**
creative	0.24		**1.13**	0.08		**0.19**	**0.34**		0.33
thrilling	0.20		**0.61**	0.06		**0.16**	0.08		**0.46**
adventurous	**0.20**		0.18	0.08		**0.11**	0.14		**0.35**
respected	0.13		**0.23**	0.05		**0.10**	0.12		**0.20**
powerful	0.09		**0.28**	0.18		**0.24**	0.13		**0.45**
superior	0.08		**0.37**	0.06		**0.21**	0.27		**0.44**
unique	**0.36**		0.14	0.07		**0.11**	0.30		**0.48**
valuable	**0.21**		0.08	0.18		**0.22**	0.22		**0.57**
likable	0.21		**0.37**	0.09		**0.10**	0.12		**0.18**

Table 3: Variances per item across brands in the three categories. For easier comparison between ratings and RT measure, higher variances are highlighted in bold. Asterisks mark a significant difference between mean variances: ** p<.01, * p<.05.

	Variability of item agreement across brands resulting from different measurement approaches:					
	INSURANCE		HAIR COLORATION		SHAMPOO	
	Variance of proportion of top-4 answers across brands	Variance of proportion of yes answers across brands	Variance of proportion of top-4 answers across brands	Variance of proportion of yes answers across brands	Variance of proportion of top-4 answers across brands	Variance of proportion of yes answers across brands
MEAN VARIANCE (attachment)	*.013*	***.013***	***.010****	> *.007*	*.010*	***.014***
affectionate	**.017**	.015	.008	.008	.007	**.018**
friendly	.005	**.006**	.006	.006	.004	**.016**
loved	.020	**.024**	**.008**	.004	**.014**	.013
peaceful	**.014**	.007	**.010**	.003	.010	**.030**
passionate	**.016**	.010	.010	**.019**	.013	**.017**
delighted	**.022**	.010	.005	**.010**	.008	**.010**
captivated	**.024**	.014	**.013**	.004	**.019**	.007
connected	.003	**.020**	**.023**	.013	**.011**	.002
bonded	.016	**.022**	**.011**	.004	**.010**	.004
attached	**.019**	.014	**.009**	.002	**.013**	.006
honest	**.009**	.007	**.007**	.003	.010	**.020**
dependable	**.009**	.008	**.010**	.005	.006	**.014**
trusting	.005	**.010**	**.010**	.005	.009	**.013**
fair	.006	**.021**	.006	.006	.007	**.030**
MEAN VARIANCE (association)	*.008*	< ***.016****	*.007*	***.008***	*.012*	< ***.015****
dutiful	.003	**.031**	.004	**.015**	.009	**.012**
disciplined	**.005**	.004	**.011**	.006	.007	**.009**
approved	.005	**.009**	**.007**	.004	.018	**.022**
reliable	.002	**.021**	**.005**	.002	**.004**	.003
easygoing	.004	**.008**	.003	**.011**	**.010**	.008
delightful	.008	**.019**	**.006**	.003	**.021**	.011
lively	**.012**	.011	.010	**.015**	.014	**.017**
creative	.018	**.052**	.005	**.012**	.017	**.020**
thrilling	.005	**.027**	.006	**.007**	.006	**.017**
adventurous	.006	**.007**	**.006**	.004	**.016**	.013
respected	.003	**.013**	.011	.011	.007	**.008**
powerful	.004	**.010**	**.012**	.010	.010	**.024**
superior	.010	**.017**	.003	**.008**	.017	**.023**
unique	**.017**	.007	**.008**	.003	.017	**.025**
valuable	**.013**	.003	.008	.008	.010	**.027**
likable	.010	**.015**	.005	**.008**	.006	**.008**

Table 4: Variances of proportion of top-4 respectively yes answers per item across brands in the three categories. For easier comparison of top-4 and yes proportions, higher variances are highlighted in bold. Asterisks mark a significant difference between mean variances: * p<.05.

In the same ways, we compute the variance of our transformed RT measure per brand across all items. Then, we compare the variances per brand of the two measures in a t-test.

On average across all brands, the variance of the RT measure is higher for both attachment and association items (independent-samples t-test, two-tailed, p<.001 for both attachment and associations; see Table 1). When looking at each brand separately, we see that the variance of the RT measure is higher than the variance of ratings for 23 out of 24 brands. This again holds for both associations and attachment items.

However, one may argue that response times show more variance simply because, as noted earlier in the paper, they are naturally very noisy and easily affected by interruption, item length, brand familiarity etc. So, an interesting question is whether the spontaneous yes-no answers are still more differentiated when we ignore the response times, compared to agreement expressed on a rating scale. We therefore also compared the variance of the percentage of yes answers across all items with the variance of the percentage of top-4 boxes of the rating scales. We thus deliberately ignore the response times to see whether the result above can be confirmed. This is indeed the case, as can be seen in Table 2. On average across all brands, the variance of the percentage of yes answers from the dichotomous answer conditions is higher than the variance of the percentage of top-4 boxes in the rating conditions (independent-samples t-test, two-tailed, p<.001 for both attachment and associations). When looking at each brand separately, we see that the variance of the percentage of yes answers is higher than the variance of the percentage of top-4 boxes for 21 out of 24 brands for the attachment items, and for 23 out of 24 brands for the associations items.

b) Variability between different brands

Second, we test whether our RT measure leads to greater differentiation between brands. We compute the mean rating of each item and explore how much the mean ratings vary across all brands of a category. This variance is compared with the variance of our transformed RT measure per item across all brands of a category.

On average, the variance of the RT measure for attachment items is significantly higher for two of the three product categories (independent-samples t-test, two-tailed, insurance: p=.004; shampoo: p=.006; see Table 3). For hair coloration, the variance of the RT measure is also higher but the difference is not significant (p=.153). For associations, the variance of RT measure is significantly higher for all three categories (independent-samples t-test, two-tailed, insurance: p=.021; hair coloration: p=.010; shampoo: p=.002). When looking at each item separately, we see that the variance of the RT measure for the attachment items is higher than the variance of rating for 12 out of 14 items in the insurance category, 8 items in the hair coloration category and 10 items in the shampoo category. When looking at the associations items, we see that the variance of the RT measure is higher than the variance of rating for 10 out of 16 items in the insurance category, 8 items in the hair coloration category and 12 items in the shampoo category.

Again, we also look at how differentiated attribute profiles are for the brands when ignoring noisy response times. For this, we compute the variance of the proportion of top-4 boxes from the ratings per item across all brands of a category and compare it with the variance of the proportion of yes answers in the dichotomous answer conditions per item across all brands of a category.

On average across all attachment items, the variance of the proportion of yes answers from the dichotomous answer condition is not significantly higher than the variance of the proportion of top-4 boxes from the rating condition (see Table 4). For hair coloration, the average between-brands variance is even smaller for the proportion of yes answers (with .007 compared to .010 for the proportion of top-4 answers, independent-samples t-test, two-tailed, insurance: p=.927; hair coloration: p=.048; shampoo: p=.160).

On average across all associations items, the variance of the percentage of yes answers from the dichotomous answer condition is higher than the variance of the proportion of top-4 boxes from the rating condition. The difference is significant for insurance and shampoo brands (independent-samples t-test, two-tailed, insurance: p=.020; hair coloration: p=.400; shampoo: p=.045).

Overall, we see that differences in variance decrease when the two approaches are compared in terms of percentage agreement. While still present when looking at differentiation between different items for a given brand, the situation is almost even when looking at differentiation between brands. But all in all, dichotomous yes-no questions with response-time measurement lead to more differentiated brand profiles and more differentiation between brands.

Evaluation of ease and fun of use by respondents

For the usability assessments that immediately followed the brand evaluations, we merely looked at differences between ratings versus yes-no questions with response-time measurement and analyzed associations and attachment items together. In terms of ease of use—ease of the task and understandability—both approaches received highly similar ratings (see Figure 3). However, yes-no questions with response-time measurement were rated as more interesting (independent-samples t-test, two-tailed, M=5.01 for RT and M=4.35 for rating, p<.001), more fun (M=5.21 for RT and M=4.51 for rating, p<.001), and less tiring (M=2.55 for RT and M=3.45 for rating, p<.001) than rating scales.

Validity: Agreement with IAT results

Since the results of the IAT are considered to represent a truly implicit attitude measure, we compared the results of the ratings and the response times to the results from the IAT. We analyzed the IAT results according to the updated guidelines in Greenwald et al.

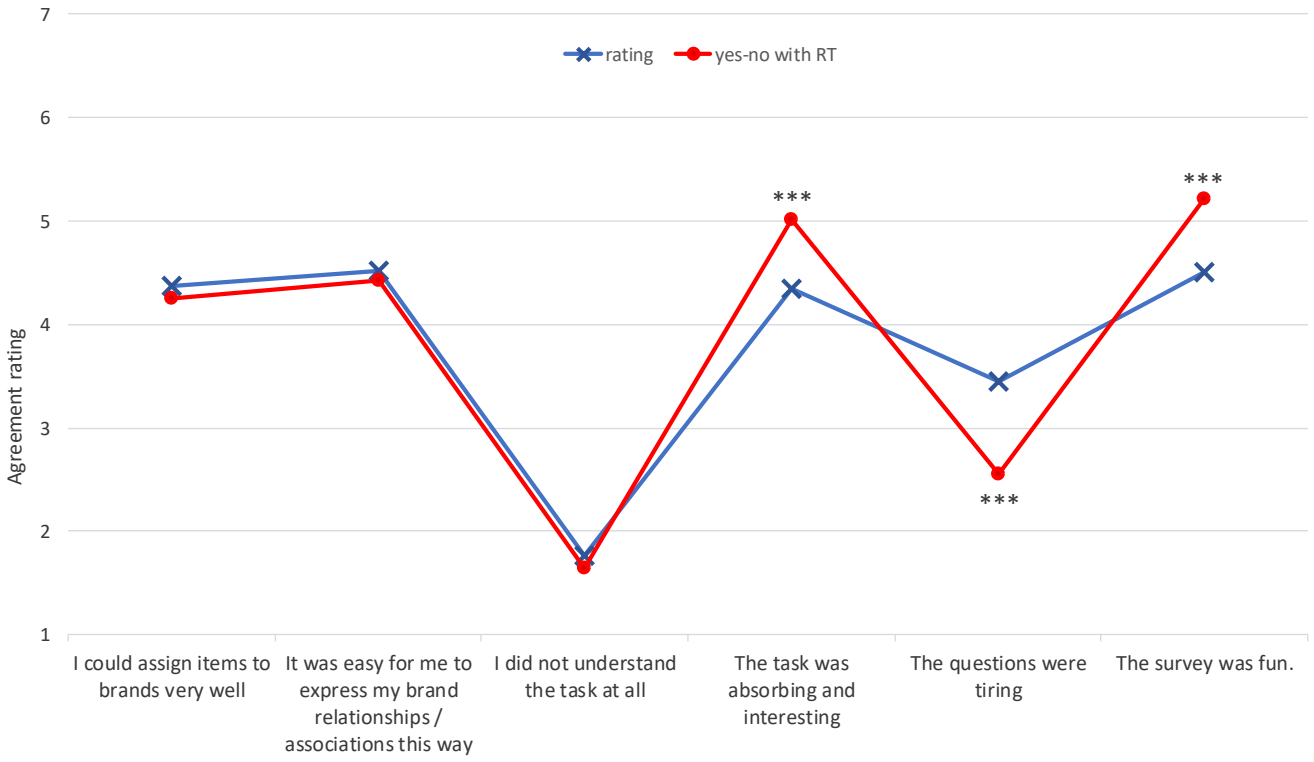

Figure 3: Mean usability assessments of ratings vs. yes-no questions with response-time measurement on six items. Asterisks mark significant differences between formats: *** p<.001.

(2003) and computed individual D scores for test and for retest. The D score represents a relative measure indicating degree of preference, or closeness of association with oneself, for one brand relative to the other. It is based on response-time differences between the combined classification tasks of brands and words, intended to identify which brand is more closely associated with positive words, or self-related words (for the detailed scoring procedure, please refer to Lane et al., 2007, Table 3.3, p. 92).

To evaluate whether our online, self-administered IAT yielded reliable results, we computed test-retest correlations for both versions of the IAT, between D scores from the first measurement point and D scores from the second measurement point one week later. Correlations were computed across product categories. For both IAT versions, a test-retest correlation of r=.52 (n=564 for the self-other IAT, n=557 for the preference IAT), respectively, was achieved, which is very close to the median correlation across 13 studies of r=.50 reported by Lane et al. (2007, p. 71).

We then explored whether the RT measure shows greater correspondence with the IAT. For this, we again focus on the first measurement point. The D score is an overall difference measure. Thus, for rating and RT measures, we first computed averages across items for each brand m_A and m_B and compute the difference $m_A - m_B$ for each respondent. Then, the correlations between these differences and the respondents' D scores are computed. Table 5 shows that correlations with D scores hardly differ between rating and RT measures.

	Correlations between brand difference scores and IAT outcomes:	
	Ratings	RT measures
Attachment	0.37	0.36
Associations	0.39	0.38

Table 5: Correlation between differences in brand rating and the IAT D score, and differences in brand RT measures and D score, respectively.

	Agreement about preferred brand:	
	Rating vs. IAT	RT measure vs. IAT
Attachment	0.66	0.62
Associations	0.64	0.61

Table 6: Proportion of respondents indicating preference for the same brand in rating and IAT, and RT measure and IAT, respectively.

	Correlations between brand difference scores and brand differences in behavioral performance indicators:			
	Ratings vs. purchase probabilities (GPC)	RT measures vs. purchase probabilities (GPC)	Ratings vs. behavioral intentions (BPI)	RT measures vs. behavioral intentions (BPI)
Attachment	0.48	0.38	0.68	0.57
Associations	0.41	0.33	0.68	0.57

Table 7: Correlation of differences in purchase probabilities from GPC and rating respectively RT-measure differences, and differences in mean BPI behavioral intention statements and rating respectively RT-measure differences.

	Agreement about preferred brand:			
	Ratings vs. purchase probabilities (GPC)	RT measures vs. purchase probabilities (GPC)	Ratings vs. behavioral intentions (BPI)	RT measures vs. behavioral intentions (BPI)
Attachment	0.73	0.63	0.77	0.68
Associations	0.62	0.61	0.74	0.72

Table 8: Proportion of respondents indicating the same preference in rating vs. GPC purchase probability, RT measure and GPC purchase probability, rating and BPI behavioral intention statements, and RT measure and BPI behavioral intention statements.

Second, we computed for how many respondents our rating and RT measures agree with D scores in terms of which brand is evaluated as superior. The respective proportions can be found in Table 6. Rating and RT measures hardly differ in terms of agreement with D score, with even slightly higher agreement between rating and IAT D score.

In sum, our results indicate that agreement of IAT with response times is not higher than with traditional rating scales. Thus, there is no indication that simple RT measures are more suitable to capture implicit aspects of brand attitudes than traditional ratings.

Validity: Agreement with GPC results and behavioral statements

To evaluate to what extent rating and RT measures are in line with other brand performance indicators, we compared them with GPC results and the behavioral statements asked towards the end of the questionnaire. First, we again focused on the differences in respondents' average brand evaluations. We correlated brand differences in ratings or RT measures, respectively, with brand differences in individual purchase probabilities derived from GPC(Wildner, 2003), and differences in mean ratings of the four BPI behavioral intentions items (Högl and Hupp, 2004). As can be seen in Table 7, the correlations with both GPC purchase probability and BPI behavioral intentions are higher for rating than for RT measure.

Second, we computed for how many respondents our rating and RT measures agree with GPC purchase probabilities and BPI behavior statements results in terms of which brand scores higher. The respective proportions can be found in Table 8. Rating results still show–sometimes slightly, but consistently–higher agreement with GPC and BPI results than RT-measure results.

Traditional ratings consistently demonstrate higher agreement than the RT measure with brand performance indicators derived from a choice based procedure and from behavioral intention statements. This is not only shown by higher correlations of the brand evaluation difference scores that may be noisier for the RT measure, but also by higher agreement in terms of preferred brand. One may argue that the brand performance indicators that were used also belong to the realm of explicit measures and may thus not be suited to evaluate the validity of a measure intended to at least partially capture implicit attitudes. We, therefore, now turn to test-retest reliability, for which the two measurement approaches are compared with themselves at two measurement points.

Test-retest reliability

As for the first questionnaire, we also remove respondents from the retest dataset whose response times in the dichotomous yes-no question conditions are below 300 ms in 20% of the items or above 3500 ms in 50% of the items or more. The remaining sample consists of n=1126 respondents who participated twice and fulfilled criteria at both measurement points.

First, we compared on an individual respondent level the rating values at t_1 and t_2 as well as the RT measures at t_1 and t_2 to see how stable the evaluations were. We focused on whether the average difference between the two assigned brands remained stable. Thus, we again used the differences in respondents' average brand evaluations and correlated differences at t_1 with differences at t_2. As can be seen in Table 9, the correlation between t_1 and t_2 is higher for the rating differences than for the RT-measure differences both for attachment and

for associations items.

	Test-retest reliabilty:	
	Ratings	RT measure
Attachment	0.78	0.65
Associations	0.78	0.62

Table 9: Correlation of brand rating differences in initial test and retest, and brand RT-measure differences in initial test and retest.

Second, acknowledging that response times have a disadvantage due to their noisy nature in terms of test-retest reliability, we also counted individual cases in which agreement remained stable. That is, for the rating scale, we counted how often a top-4 box remained a top-4 box answer or a bottom-3 box remained a bottom-3 box answer. For the response condition, we counted how often a yes response was repeated and how often a no response was repeated. The respective proportions of stable answers are reported in Table 10. Rating answers, thus, show higher stability even when compared with the proportion of yes answers, ignoring noisy response times.

	Proportion of identical answers at t_1 and t_2:	
	Rating condition	Yes-no condition
Attachment	0.80	0.73
Associations	0.81	0.75

Table 10: Proportion of identical answers in initial test and retest with respect to ratings (either top-4 or bottom-3 in both waves) and dichotomous answers (either yes or no in both waves).

5 Discussion

The results of the comparisons between rating and RT measure show a clear picture. While the simple RT measure produces higher variance between brands and items, rating scales show higher test-retest reliability and higher correlation with purchase likelihood, willingness to recommend, and purchase intent. Moreover, the simple RT measure does not show higher agreement with the IAT compared to rating scales. Thus, the alleged higher differentiation that is suggested by the RT measure is attributed to greater error variance rather than reliable attitude differences. Merely in terms of fun of use the simple RT measures performed better than rating scales.

Surprisingly, even when response times are ignored and data analysis is focused on proportion of yes answers, results tend to be inferior compared to ratings: In our results, also the proportion of yes answers achieved inferior results in terms of validity and reliability compared to the percentage of top-4 boxes.

It should be noted, however, that the approach we took is not suited to raise doubts about the general validity of response times as a measure of attitude strength. Our purpose was to thoroughly test an approach that is frequently applied in consumer research practice (e.g., Heinsen and Lorenz, 2011; Scheier, 2006; Schmidt et al., 2015), probably for efficiency reasons, to present single items with simple yes-no answers and analyze time to respond. In contrast, validity and reliability of several truly implicit tests, in which the question purpose is not evident to respondents, are well-documented, including their limitations (e.g., Bosson et al., 2000; de Houwer, 2008; Krause et al., 2011). And even direct questions with response-time measurement can reveal associative strength between concepts (cf., e.g., Fazio, 1990; Sternberg, 2010). To be informative, however, usually several trials per concept are required so response times can be averaged, attenuating the effect of noise. Also, it is recommended to control for factors affecting response times, such as word length, word familiarity, order effects, and individual differences. If these recommendations are followed, direct response-time measures of attitude strength will require more of respondents' time and more effort in data analysis, but reliability and validity are likely to increase. For response-time approaches to really add value to brand attitude measurement and provide reliable insights for managers, it is inevitable that such recommended precautions are empirically evaluated in future studies.

6 Conclusion

Our results show that simple response-time measures are neither a valid alternative to rating scales nor an efficient substitute for sophisticated implicit methods such as the IAT. They suffer from inferior validity and reliability, while at the same time not showing higher agreement with full-blown IAT results compared to traditional rating scales. The expenditure associated with the IAT, in turn, seems only justified for highly sensitive topics. At the same time, there is room for improvement for standard rating scales. Improved handling via drag and drop and more intuitive pictorial designs that dispense with numbers may represent promising developments for maintaining respondent motivation.

References

Aaker, J. (1997). Dimensions of brand personality. *Journal of Marketing Research*, 34(3):347–356.

Asendorpf, J. and Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, 83(2):380–393.

Bischof, N. (1993). Untersuchungen zur Systemanalyse der sozialen Motivation I: Die Regulation der sozialen

Distanz âĂŞ Von der Feldtheorie zur Systemtheorie [On the regulation of social distance âĂŞ from field theory to systems theory]. *Zeitschrift für Psychologie*, 201:5–43.

Bosson, J., Swann, W., and Pennebaker, J. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79(4):631–643.

Bridger, D. (2015). *Decoding the irrational consumer: How to commission, run and generate insights from neuromarketing research*. Kogan Page, London.

Brunel, F., Tietje, B., and Greenwald, A. (2004). Is the implicit association test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology*, 14(4):385–404.

Cho, E., Fiore, A., and Russell, D. (2015). Validation of a fashion brand image scale capturing cognitive, sensory, and affective associations: Testing its role in an extended brand equity model. *Psychology & Marketing*, 32(1):28–48.

Cunningham, W., Johnson, M., Gatenby, J., and Gore, J. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology*, 85(4):639–649.

Cunningham, W., Johnson, M., Gatenby, J., Gore, J., and Banaji, M. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, 15(12):806–813.

de Houwer, J. (2008). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In Petty, R., Fazio, R., and Brinol, P., editors, *Attitudes: Insights from the new implicit measures*, pages 361–390. Lawrence Erlbaum, Mahwah, NJ.

Fazio, R. (1990). A practical guide to the use of response latency in social psychological research. In Hendrick, C. and Clark, M., editors, *Review of personality and social psychology, Vol. 11. Research methods in personality and social psychology*, pages 74–97. Sage Publications, Inc., Thousand Oaks, CA.

Fazio, R. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15(2):115–141.

Fournier, S. (1998). Consumers and their brands: Developing relationship theory in consumer research. *Journal of Consumer Research*, 24(4):343–353.

Fournier, S. (2009). Lessons learned about consumers' relationships with their brands. In MacInnis, D., Park, C., and Priester, J., editors, *Handbook of brand relationships*, pages 5–23. M E Sharpe, Armonk, NY.

Gaspar, C. (2011). Vertrauen will verdient sein. In und GfK Panel Services Deutschland, G. V., editor, *30. Unternehmergespräch Kronberg. Erfolgsfaktor Vertrauen. Wie Markenvertrauen entsteht und wirkt*, pages 63–77. Nürnberg.

Gattol, V., Sääksjärvi, M., and Carbon, C.-C. (2011). Extending the Implicit Association Test (IAT): Assessing consumer attitudes based on multi-dimensional implicit associations. *PLoS ONE*, 6(1):e15849.

Gawronski, B. and Conrey, F. (2004). Der Implizite Assoziationstest als Maß automatisch aktivierter Assoziationen: Reichweite und Grenzen. *Psychologische Rundschau*, 55(3):118–126.

Glashouwer, K., Smulders, F., de Jong, P., Roefs, A., and Wiers, R. (2013). Measuring automatic associations: Validation of algorithms for the Implicit Association Test (IAT) in a laboratory setting. *Journal of Behavior Therapy and Experimental Psychiatry*, 44(1):105–113.

Greenwald, A., Banaji, M., Rudman, L., Farnham, S., Nosek, B., and Mellott, D. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109(1):3–29.

Greenwald, A., McGhee, D., and Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6):1464–1480.

Greenwald, A., Nosek, B., and Banaji, M. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2):197.

Heinsen, S. and Lorenz, A. (2011). Implizite Messverfahren – die Abkürzung zum Unbewussten. In und Sozialforschung, V. S. M., editor, *Jahrbuch 2011 Markt- und Sozialforschung*, pages 56–58. Geschäftsstelle vsms Verband Schweizer Markt- und Sozialforschung, Zürich.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31(10):1369–1385.

Högl, S. and Hupp, O. (2004). Brand Performance Measurement mit dem Brand ASsessment System (BASS). In Schirmansky, A., editor, *Der Wert der Marke: Markenbewertungsverfahren für ein erfolgreiches Markenmanagement*, pages 124–145. Vahlen, München.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, NY.

Krause, S., Back, M., Egloff, B., and Schmukle, S. (2011). Reliability of implicit self-esteem measures revisited. *European Journal of Personality*, 25(3):239–251.

Lane, K., Banaji, M., Nosek, B., and Greenwald, A. (2007). Understanding and using the Implicit Association Test: IV – What we know (so far) about the method. In Wittenbrink, B. and Schwarz, N., editors, *Implicit Measures of Attitudes*, pages 59–102. The Guilford Press, New York, NY.

Mayerl, J. and Urban, D. (2008). *Antwortreaktionszeiten in Survey-Analysen: Messung, Auswertung und Anwendungen*. VS Verlag für Sozialwissenschaften, Wiesbaden.

McConnell, A. and Leibold, J. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5):435–442.

Nosek, B., Bar-Anan, Y., Sriram, N., Axt, J., and Greenwald, A. (2014). Understanding and Using the Brief Implicit Association Test: Recommended Scoring Procedures. *PLoS ONE*, 9(12):1–31.

Nosek, B., Greenwald, A., and Banaji, M. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In Bargh, J. A., editor, *Automatic Processes in Social Thinking and Behavior*, pages 265–292. Psychology Press, New York, NY.

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6(1):101–115.

Peterson, R. and Wilson, W. (1992). Measuring customer satisfaction: Fact and artifact. *Journal of applied psychology*, 20(1):69.

Petty, R. and Cacioppo, J. (1986). The elaboration likelihood model of persuasion. In Berkowitz, L., editor, *Advances in experimental social psychology, Vol. 19*, pages 123–205. Academic Press, New York, NY.

Scheier, C. (2006). Das Unbewusste messbar machen. *Absatzwirtschaft*, 10:42–45.

Scheier, C. (2008). Neuromarketing – über den Mehrwert der Hirnforschung für das Marketing . In Kreutzer, R. and Merkle, W., editors, *Die neue Macht des Marketing – Marketing zwischen Emotion, Innovation und Präzision*, pages 305–323. Gabler, Wiesbaden.

Scheier, C. and Held, D. (2006). *Wie Werbung wirkt: Erkenntnisse des Neuromarketing*. Haufe, Planegg.

Scheier, C. and Held, D. (2007). Die Neuentdeckung des Unbewussten: Was wir von der Hirnforschung für Markenführung und Marktforschung lernen können. In Meyer, H., editor, *Marken-Management 2008/2009. Jahrbuch für Strategie und Praxis der Markenführung*, pages 229–255. Deutscher Fachverlag, Frankfurt am Main.

Scheier, C. and Held, D. (2010). Markenmanagement nachhaltig implementieren. Erkenntnisse des Neuromarketing. In Bruhn, M. and Köhler, R., editors, *Wie Marken wirken. Impulse aus der Neuroökonomie für die Markenführung*, pages 200–214. Vahlen, München.

Schmidt, S., Langner, S., Schießl, M., and Wiedmann, K. (2015). Touchpoint research and the moment of truth. In Schmidt, H. and Baumgarth, C., editors, *Forum Markenforschung*, pages 153–165. Springer Gabler, Wiesbaden.

Shuv-Ami, A. (2012). Brand commitment: A new four-dimensional (4 Es) conceptualisation and scale. *Journal of Customer Behaviour*, 11(3):281–305.

Sternberg, S. (2010). *Reaction-time experimentation*. Unpublished manuscript retrieved from http://www.psych.upenn.edu/~saul/ [March 08, 2018].

Strack, F. and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Peronality and Social Psychology Review*, 8(3):220–247.

The Economist (2013). *Nothing more than feelings*. Retrieved from https://www.economist.com/news/business/21591165-admen-have-made-marketing-guru-daniel-kahneman-prizewinning-psychologist-nothing-more.

Thomson, M., MacInnis, D. J., and Park, C. (2005). The ties that bind: Measuring the strength of consumers' emotional attachments to brands. *Journal of Consumer Psychology*, 15(1):77–91.

Thorndike, E. (1920). A constant error in psychological ratings. *Journal of applied psychology*, 4(1):25–29.

Wildner, R. (2003). Marktforschung für den Preis. *Jahrbuch der Absatz- und Verbrauchsforschung*, 49(1):4–26.