BART DE LANGHE, STEFANO PUNTONI, DANIEL FERNANDES, and STIJN M.J. VAN OSSELAER*

In an increasingly globalized marketplace, it is common for marketing researchers to collect data from respondents who are not native speakers of the language in which the questions are formulated. Examples include online customer ratings and internal marketing initiatives in multinational corporations. This raises the issue of whether providing responses on rating scales in a person's native versus second language exerts a systematic influence on the responses obtained. This article documents the anchor contraction effect (ACE), the systematic tendency to report more intense emotions when answering questions using rating scales in a nonnative language than in the native language. Nine studies (1) establish ACE, test the underlying process, and rule out alternative explanations; (2) examine the generalizability of ACE across a range of situations, measures, and response scale formats; and (3) explore managerially relevant and easily implementable corrective techniques.

*Keywords*: international marketing research, language, bilingualism, emotions

# The Anchor Contraction Effect in International Marketing Research

Technological, social, and economic changes are linking an ever-growing number of people around the world into complex patterns of interdependence (Dicken 2007). Aspects of globalization such as the growth of the Internet, the cosmopolitanism of large cities, and cross-national trade imply that, compared with a few decades ago, a much larger share of marketing research data is now collected from multilingual or multicultural respondents. Although marketing research agencies often translate surveys into respondents' native language, there are many cases in which data are collected in a respondent's second language, typically English. This raises the issue of whether providing responses on rating scales in a person's native (henceforth, L1) versus second language (L2) exerts a systematic influence on the responses obtained. In this article, we document the anchor contraction effect (ACE) for bilinguals' L2, which is the systematic tendency for people to report more intense emotions when answering questions using L2 rating scales than when using L1 rating scales. To emphasize the substantive importance of this issue, consider the following situations:

- Online customer ratings are an increasingly important and visible feature of online retailers (Chevalier and Mayzlin 2006). For example, Amazon.com offers customers the opportunity to rate any product using the emotional statements "I hate it" and "I love it." Regardless of their native language, consumers around the world contribute ratings to the Web site, and language (L1 vs. L2) might exert a significant influence on such ratings.
- Many societies are becoming increasingly multilingual. For example, by the year 2025, more than half of all U.S. families with children will be multicultural (Anderson 2009). For governments and firms, this trend raises the issue of how to interpret answers to questions expressed in the country's official language by ethnic minorities who are nonnative speakers of that language (Richard and Toffoli 2009). For example, in a study by the Pew Research Center for the People and the Press

*Bart de Langhe is a doctoral student (e-mail: blanghe@rsm.nl), Stefano Puntoni is Associate Professor of Marketing (e-mail: spuntoni@rsm.nl), Daniel Fernandes is a doctoral student (e-mail: dfernandes@rsm.nl), and Stijn M.J. van Osselaer is Professor of Marketing (e-mail: sosselaer@rsm.nl), Rotterdam School of Management, Erasmus University, the Netherlands. Bart de Langhe will join the Leeds School of Business, University of Colorado at Boulder, as Assistant Professor of Marketing. The first two authors contributed equally to the article. The authors thank participants of the 2009 Kortenberg seminar at the Catholic University of Leuven, Martijn de Jong, and Dirk Smeesters for their comments. The financial assistance of the Erasmus Research Institute of Management is gratefully acknowledged. Jan-Benedict Steenkamp served as associate editor for this article.


ISSN: 0022-2437 (print), 1547-7193 (electronic)

*Journal of Marketing Research*
Vol. XLVIII (April 2011), 366–380

(2001) conducted in the United States, Hispanic respondents, many of whom are not native English speakers, expressed more worry about new terrorist attacks than other respondents. Because the survey was in English for all participants, the interpretation of this finding is complicated by the notion that we do not know whether the finding is due to differences in perceptions or to language.

- Another context in which ACE is likely to play a role is customer satisfaction measures in hotels and other sites commonly visited by international travelers. For example, the emotional anchoring point "happy" is often used in customer satisfaction measures (e.g., Bruner, James, and Hensel 2001). In such cases, ACE could lead foreign visitors to express more positive opinions than local residents and possibly mislead managers.
- Multinational corporations that operate across a large number of countries often adopt English as their official language (Marschan-Piekkari, Welch, and Welch 1999). Consider a hypothetical situation in which a multinational firm conducts a survey of its employees to assess the prevalence of work-related emotional distress and anxiety. The firm's official language is English, the language in which the survey is expressed. Employees from both British and Spanish subsidiaries complete the survey, with the consequence that some respondents answer emotional items using the L1 rating scale while others use the L2 rating scale.

In these scenarios, as in many others, bilingual people answer questions that probe emotional processes using L2 anchoring points. These responses are then compared with, or averaged with, the responses of people who answered the same questions in L1. In these situations, ACE can lead to inflated error terms or, worse, to wrong inferences. For example, consider again the last scenario. Upon completion of the employee satisfaction survey in the multinational firm, if the results show that Spanish employees tend to report more intense aversive emotional states than British employees, management may decide to transfer resources from the personnel department of the British subsidiary to that of the Spanish subsidiary. However, the importance of ACE in this situation indicates that differences in the answers of Spanish and British employees may not be due to any meaningful disparity in working conditions but rather to the notion that, unlike their British colleagues, Spanish respondents answered the survey using L2 rating scales.

In the next section, we review relevant literature on measurement and bilingualism to derive the central prediction that L2 rating scales yield more extreme responses than L1 rating scales in the case of emotion-laden items. We then present the findings of nine experimental studies with a total of more than 1000 respondents. The experimental approach enables us to (1) establish ACE, test the underlying process, and rule out alternative explanations; (2) examine the generalizability of ACE across a range of situations, measures, and response scale formats; and (3) explore managerially relevant and easily implementable corrective techniques.

## *THEORETICAL BACKGROUND*

Previous research has described several important factors that can explain higher or lower ratings in surveys. A large body of research has studied response styles, which reflect a person's tendency to systematically respond to questions on some basis other than what the questions were designed to measure (Baumgartner and Steenkamp 2001). For example, people differ in their tendency to give extreme responses (Greenleaf 1992; Van Rosmalen, Van Herk, and Groenen 2010). In addition to individual differences, responses may also be systematically affected by survey characteristics. For example, De Jong et al. (2008) show that extreme responding is influenced by the length of the questions. In addition, the meaning and location of verbal labels and numeric values accompanying rating scales might cause differential responses to the scales (Baumgartner and Steenkamp 2001; Ostrom 1966; Schwarz et al. 1991; Upshaw 1965; Wildt and Mazis 1978). In particular, stimulus ratings are a function of the interpretation of the scale's anchoring points (Ostrom 1966; Upshaw 1965). All else being equal, more intense verbal labels on the anchoring points of a rating scale lead respondents to move away from the ends of the scale. For example, the same level of experienced happiness should lead people to select a lower number on a seven-point unipolar scale when the extreme anchor is worded as "ecstatic" or "overjoyed" than when it is worded as "pleased" or "glad." The reason for this is that the same experienced happiness is lower relative to the more extreme anchor (e.g., "ecstatic") than to the less extreme anchor (e.g., "pleased").

Recent research has shown that bilinguals tend to experience L1 words as more emotionally intense than the same words in L2 (Altarriba 2003; Harris, Gleason, and Aycicegi 2006; Pavlenko 2005; Puntoni, De Langhe, and Van Osselaer 2009). This literature attributes a key role to autobiographical memories in determining the emotional intensity of words (Marian and Kaushanskaya 2004). Because of the language specificity of autobiographical memory (Marian and Neisser 2000), L1 (L2) words automatically trigger an emotional echo from previous L1 (L2) experiences (Puntoni, De Langhe, and Van Osselaer 2009). Because autobiographical memories in a person's native language are typically both more frequent (Puntoni, De Langhe, and Van Osselaer 2009) and more emotional (Harris, Gleason, and Aycicegi 2006) than L2 autobiographical memories, L1 stimuli tend to elicit more intense emotional experiences. For example, Puntoni, De Langhe, and Van Osselaer (2009) find that advertising slogans expressed in L1 generate stronger emotional reactions than those expressed in L2. This stream of research establishes the influence of the language of the to-be-rated target on its perceived emotional intensity.

Traditional measurement theory posits that an observed score is the result of a true score plus a measurement error. Although previous research on bilingualism demonstrates the influence of the language of the target on the true score (e.g., Harris, Aycicegi, and Gleason 2003; Puntoni, De Langhe, and Van Osselaer 2009), no research has examined whether the language of the measurement instrument introduces a systematic component into the error term. However, it is likely that, in addition to to-be-rated words, emotion-related words used to anchor a scale might be experienced as less intense when they are written in L2 than when they are written in L1. For example, nonnative English speakers might experience emotional scale anchors such as "happy" or "sad" as *less* intense in English than in their native language. If to-be-rated stimuli are then judged *relative* to those (less vs. more intensely experienced) anchors, we should find *more* extreme (i.e., less neutral) answers if the emotional scale anchors are in L2 than in L1. That is, the nonnative English speaker should judge the same stimulus

as more intense in comparison to the less intensely experienced English scale anchor (e.g., "happy") than in comparison to its more intensely experienced L1 equivalent. In summary, we predict that the same stimuli will be rated as emotionally less intense when the scale anchors are labeled in L1 than in L2. Stated differently, emotional anchoring points presented in L2 should contract the scale relative to anchoring points presented in L1. Figure 1 presents a schematic representation of ACE in the case of native speakers of Dutch confronted with a survey in English.

## *STUDY 1*

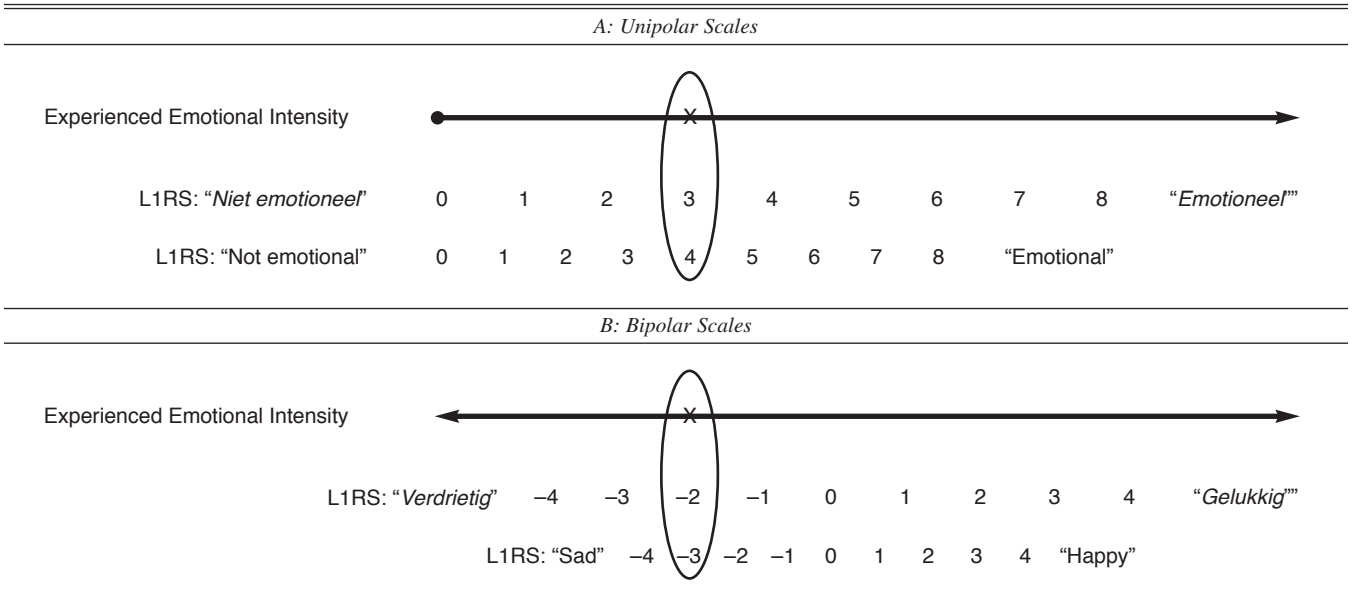
We designed Study 1 to provide an initial demonstration of ACE in a context that controls for the possible effect of language stereotypes (Leclerc, Schmitt, and Dube 1994). If cultures (and associated languages) vary in their perceived or actual warmth and emotional expressiveness (Cuddy et al. 2009), emotional anchoring points in languages associated with more emotionally expressive cultures may be perceived as more intense. For example, it is possible that, because of sociolinguistic processes, unipolar rating scales in French (a language associated with a stereotypically hedonistic or expressive culture) tend to yield lower scores than rating scales in German (a language with a less emotional stereotype; Cuddy et al. 2009). To preclude this alternative explanation, Study 1 adopts a counterbalanced bilingual design that varies across participants which of two target languages is L1 and L2. In particular, we selected the target languages such that for half of the respondents, stereotype effects conflict with ACE, and for the other half these influences are in the same direction. In a taste test, we asked speakers of French and Dutch to evaluate a chocolate on several emotional dimensions using either French or Dutch unipolar rating scales. For half the participants Dutch was L1, and for the other half French was L1. French has been used in previous research to generate hedonic associations (Leclerc, Schmitt, and Dube 1994). In comparison, Germanic languages and associated cultures such as Dutch are typically considered less emotional (Cuddy et al. 2009). To rule out an alternative explanation for ACE based on sociolinguistic processes, we should observe a main effect of language, such that L2 rating scales lead to higher ratings of emotional intensity than L1 ratings, regardless of which of the two target languages (the stereotypically warmer or the stereotypically colder language) is L1 and L2.

### *Method*

*Design and participants*. The study used a 2 (language of the rating scales: L1 vs. L2) × 2 (native language: French vs. Dutch) between-subjects design. We collected data in Brussels, the bilingual capital city of Belgium, and recruited participants in cafeterias on the campuses of Dutch- and French-speaking sister universities. We addressed all participants in English to avoid asymmetric influences of the language of experimenter–respondent interaction. Participants included 120 proficient Dutch–English–French trilinguals (61 native Dutch speakers: $M_{age}$ = 22.74, SD = 5.58; 56 women).

*Procedure*. We invited the respondents to participate in an international research project on the taste of chocolate. The experimenter asked the participants to taste a chocolate, after which the participants completed several questions using seven-point unipolar scales (numbered from 1 to 7). We first created the booklet in English, and native Dutch and French speakers subsequently translated it. Two inde-

Figure 1
SCHEMATIC REPRESENTATION OF THE ANCHOR CONTRACTION EFFECT FOR NATIVE DUTCH SPEAKERS ANSWERING QUESTIONS IN ENGLISH




Notes: L1RS = L1 rating scales, and L2RS = L2 rating scales. (A) If respondents perceive the L1 label (e.g., "*emotioneel*") as more emotional than the corresponding L2 label (e.g., "emotional"), then for the same experienced emotional intensity evoked by the target (in the figure, "x"), the score in the L1 rating scales condition (e.g., L1RS = 3) should be lower (i.e., less extreme) than the score in the L2 rating scales condition (e.g., L2RS = 4). (B) If respondents perceive the L1 negative label (e.g., "*verdrietig*") as more intense than the corresponding L2 label (e.g., "sad"), then for the same experienced sadness ("x"), the score in the L1 rating scales condition (e.g., L1RS = –2) should be higher (i.e., less extreme) than the score in the L2 rating scales condition (e.g., L2RS = –3).

pendent judges then checked the translated booklets for consistency with the English version. Five questions measured participants' emotional responses. First, participants rated happiness, joy, and excitement ("The taste of this chocolate makes me feel ..."; anchoring points were "not at all happy/very happy," "not at all joyful/very joyful," and "not at all excited/very excited"). Next, participants rated surprise and emotionality ("The taste of this chocolate is ..."; anchoring points were "not at all surprising/very surprising" and "not at all emotional/very emotional"; for the five items, $\alpha = .78$). Finally, participants provided basic demographic information, including their native language.

### *Results*

A repeated measures analysis of variance (ANOVA) with the five items as repeated measures and language of the rating scales and native language as between-subjects factors yielded a significant main effect of language of the rating scales ($F(1, 116) = 4.08$, $p < .05$, Cohen's $d = .37$).[1] Demonstrating ACE, participants reported more intense emotional experiences when using L2 (M = 3.36) than when using L1 rating scales (M = 3.02; for means by emotion, see Table 1). Critically, this effect was not qualified by a two-way interaction between language of the rating scales and native language ($p > .66$). In other words, the magnitude of ACE did not depend on whether the native language of the respondent was French or Dutch. (The main effect of native language was also nonsignificant; $p > .72$.[2])

### *Discussion*

Study 1 demonstrates ACE using a balanced bilingual design. Participants rated the intensity of the emotional reactions generated by tasting a chocolate higher when using L2 than when using L1 rating scales. This effect was not qualified by an interaction with native language (French vs. Dutch). Although we do not contend that language stereotypes can *never* exert an influence, this study demonstrates that they alone do not explain our results. In addition, the balanced bilingual design in Study 1 shows that ACE cannot be attributed to the selection of nonequivalent words in the L1 versus L2 conditions (i.e., picking inherently more extreme anchor words in L2 than in L1).

## *STUDY 2*

The goals of Study 2 were to replicate ACE in a different setting, with a wider array of emotions, and to test whether ACE occurs similarly for positive and negative emotions. We asked Dutch respondents to rate the intensity of five positive and five negative emotions portrayed in an animated movie using either L1 (Dutch) or L2 (English) rating scales. To control for the possible influence of the language of the target to be evaluated, we used a language-free movie. We predict higher emotional intensity scores when the movie is rated using L2 than when using L1 rating scales for both positive and negative emotions.

### *Method*

We used a $2 \times 2$ mixed design and manipulated the language of the rating scales (L1 vs. L2) between subjects and valence (positive vs. negative) within subject. Sixty-one students at a large Dutch university who were proficient English speakers ($M_{age} = 20.23$, SD = 2.41; 19 women) participated for extra course credit. All were enrolled in degree programs partially or entirely taught in English. (We used the same population for Studies 3–9.) Participants first watched a short animated movie (Pixar's *Presto*, approximately five minutes long) and then rated the intensity of ten emotions (five positive and five negative) portrayed in the movie using seven-point unipolar scales (numbered from 0 to 6) either in Dutch (L1) or English (L2). The beginning and end of the movie had been edited to hide any textual information. We selected the target emotions according to their relevance to the story and presented them in random order. The negative emotions were fear, frustration, hate, sadness, and shame. The positive emotions were happiness, hope, love, pride, and surprise.

### *Results*

A repeated measures ANOVA with the average emotional intensity of the five positive and five negative emotions as repeated measures and language as a between-subjects factor revealed a main effect of language ($F(1, 59) = 6.96$, $p = .01$, $d = .68$). Replicating ACE, participants reported more intense scores when rating the movie using L2 (M = 3.95, SD = .72) than when using L1 (M = 3.43, SD = .79) rating

[1] In all studies, to calculate effect sizes, we used Cohen's d for between-group comparisons, $\eta_p^2$ for dependent samples, and Cohen's $f^2$ for continuous predictors (e.g., Meyers-Levy, Zhu, and Jiang 2010). We manipulated the language of the rating scales between subjects in all studies; thus, Cohen's d is the measure of effect size for ACE in this article. According to Cohen (1992), d-values of .20, .50, and .80 represent small, medium, and large effect sizes, respectively.

[2] Other theoretically uninteresting effects involving the repeated emotion factor were significant—for example, the main effect of emotion ($F(4, 464) = 44.04$, $p < .0001$) and the emotion by native language interaction ($F(4, 464) = 3.21$, $p < .05$). The effect of language of the rating scales was in the predicted direction for all items (see Table 1).

Table 1
CELL MEANS (AND STANDARD DEVIATIONS) IN STUDY 1

| *Language of the Rating Scales* | *Native Language* | *Item* | | | | |
|---|---|---|---|---|---|---|
| | | *Happy* | *Joyful* | *Excited* | *Surprising* | *Emotional* |
| L2 | Dutch | 3.72 | 3.69 | 3.61 | 3.00 | 2.47 |
| | | (1.03) | (1.01) | (1.20) | (1.47) | (1.44) |
| L1 | Dutch | 3.52 | 3.68 | 3.52 | 2.80 | 1.64 |
| | | (1.08) | (1.03) | (1.04) | (1.35) | (1.19) |
| L2 | French | 3.97 | 3.67 | 3.47 | 3.03 | 3.03 |
| | | (1.38) | (1.18) | (1.41) | (1.45) | (1.35) |
| L1 | French | 3.65 | 3.37 | 3.23 | 2.65 | 2.17 |
| | | (1.11) | (1.32) | (1.33) | (1.37) | (1.07) |

scales. This main effect was not qualified by an interaction between emotion valence and language ($p > .46$), indicating that the effect of language was in the same direction for positive ($M_{L2} = 3.96$, $SD_{L2} = .86$, and $M_{L1} = 3.55$, $SD_{L1} = .81$) and negative emotions ($M_{L2} = 3.93$, $SD_{L2} = .87$, and $M_{L1} = 3.32$, $SD_{L1} = 1.15$; see Figure 2, Panel A).

*Discussion*

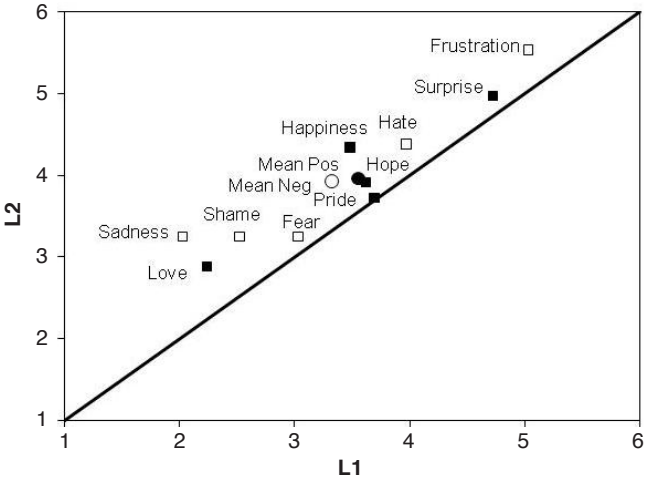
Study 2 replicates ACE with a wider set of emotions, both positive and negative. We used English as L2 because English is the language most likely to be implicated in situations in which ACE might represent an issue for marketing researchers. Greater similarity between target languages reduces the magnitude of language effects (Sunderman and Kroll 2006), and with the exception of Frisian, Dutch is the language closest to English (Finegan 1987). Thus, using Dutch as L1 constitutes a conservative test of ACE. Because participants rated both positive and negative emotions as more intense, Study 2 also shows that ACE is independent of valence.

It is worth noting that ACE is unlikely to result from a lack of language proficiency in this sample. The emotional anchors were all single words that the participants could easily translate. In addition, the effect was not weaker for those emotional anchors that are cognates (i.e., frustration–*frustratie*, hate–*haat*, hope–*hoop*) than for anchors that sound different in Dutch and English and thus are relatively more difficult to process (Costa, Caramazza, and Sebastián-Gallés 2000). Other findings throughout this article are also irreconcilable with an explanation of ACE in terms of language comprehension.

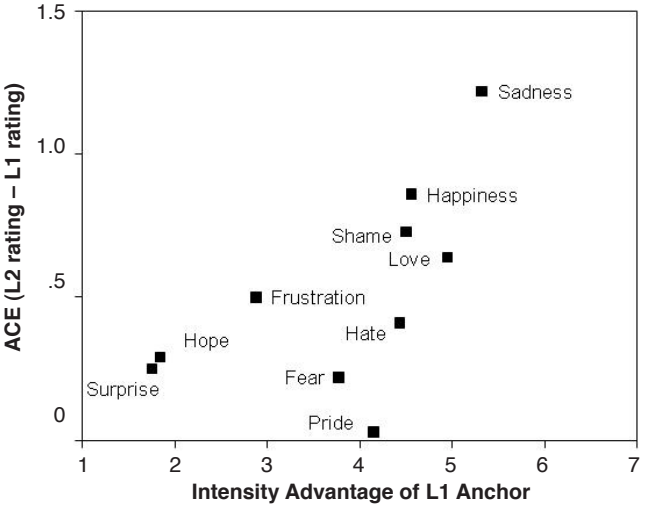
Instead, we argue that ACE is driven by a difference in perceived emotional intensity of L1 and L2 rating scale anchors. Because bilinguals perceive emotional anchoring points as more intense in L1 than in L2, the same emotional experience will be expressed with a less extreme rating in L1 than in L2. To provide initial evidence for this process, we invited 88 additional participants to rate the difference in emotional intensity between the L1 and L2 verbal representation of the ten emotions rated in Study 2 (e.g., hate–*haat*). Participants expressed ratings on a scale ranging from 0 ("no difference in emotional intensity at all") to 10 ("much stronger emotional intensity in Dutch"). Because of the small number of emotions considered in Study 2 (N = 10), we used the nonparametric Spearman rank-order correlation to assess the relationship between the L1–L2 intensity difference of the emotional anchors and the magnitude of ACE observed in Study 2. Consistent with our reasoning, ACE was larger for emotions with a larger L1–L2 intensity difference (Spearman's rho = .73, $p < .05$; see Figure 2, Panel B).

Figure 2
RESULTS OF STUDY 2

*A: ACE for Ten Specific Emotions*




*B: ACE by the Relative Intensity Advantage of the L1 Emotional Anchor*




## *STUDY 3*

Study 2 provides correlational evidence for the mechanism underlying ACE. Emotions rated as relatively more intense in L1 than in L2 tended to display larger ACE. The goal of Study 3 was to provide further evidence for the role of the perceived emotional intensity of L1 versus L2 scale anchors. Specifically, the study tests whether the perceived emotional intensity of the anchoring points mediates the effect of language on ratings of a target object. A sample of Dutch participants first rated the emotional intensity of the expression "feeling happy" in Dutch (L1) and English (L2). Subsequently, we asked the participants to rate the extent to which a specific advertisement triggered feelings of happiness using either an L1 or an L2 rating scale.

*Method*

We recruited 112 college students on campus to participate in two allegedly unrelated studies in return for a small reward ($M_{age} = 21.39$, SD = 3.03; 58 women). We randomly assigned participants to one of two language conditions (language of the booklet: L1 vs. L2). For each participant, all materials were in one language (Dutch or English), with the exception that each participant rated the emotional intensity of "feeling happy" (the mediator) in both L1 and L2.

In the first part of the study, we informed participants that the expression "feeling happy" might be perceived as more or less intense depending on the language in which it is

expressed (Dutch or English). We then asked all participants to rate the intensity of both "feeling happy" and its L1 equivalent ("*gelukkig voelen*"). For each of the two items separately, we presented participants with a right-pointing arrow next to the to-be-rated expression marked with the words "stronger emotional experience." To simplify coding, we divided the continuous line into 43 equally spaced sections. We then randomized the order of the two items.

Next, we introduced an allegedly unrelated study and informed participants that advertisers often use images to deliver emotional messages and that the study investigated the effectiveness of images in conveying emotions. We asked participants to look at an image from an actual advertisement and indicate the extent to which the image made them feel happy. We removed all textual cues from the advertisement. Participants rated the image either in L1 or in L2 on a seven-point unipolar scale ranging from 0 ("not at all happy") to 6 ("very happy").

*Results and Discussion*

To test for mediation, we estimated three regressions (Baron and Kenny 1986). First, a regression of the ad ratings on language of the rating scale (entered as a dummy variable) yielded a significant effect of language (b = .83, t(110) = 3.34, $p < .01$, d = .63). Replicating ACE, happiness ratings were higher with L2 (M = 3.60, SD = 1.31) than with L1 rating scales (M = 2.76, SD = 1.33).

Next, we estimated a regression of the perceived emotional intensity of "feeling happy" (the mediator) on language. Because we could not use scale anchors when measuring the intensity of the scale anchors in L1 and L2, we had to control for heterogeneity across participants in the interpretation of the continuous-arrow response scale. For this purpose, we standardized perceived emotional intensity by dividing each participant's rating of "feeling happy" or "*gelukkig voelen*" (depending on the experimental condition) by the average of the participant's ratings of "feeling happy" and "*gelukkig voelen*." Replicating the findings of Puntoni, De Langhe, and Van Osselaer (2009), we found a significant effect of language (b = –.11, t(110) = –3.66, $p < .001$, d = –.69): Specifically, participants rated "feeling happy" (M = .94, SD = .16) as less intense than "*gelukkig voelen*" (M = 1.05, SD = .16).

Finally, a third model regressed the ad ratings on the language of the rating scale and the perceived emotional intensity of the scale anchor. This model yielded a significant effect of the scale anchor's perceived emotional intensity (b = –1.59, t(109) = –2.06, $p < .05$, Cohen's $f^2 = .06$), such that when the anchor was perceived as more emotionally intense, ad ratings were lower. Moreover, although the effect of rating scale language on the ad ratings remained significant (b = .66, t(109) = 2.52, $p < .05$), the addition of the perceived emotional intensity of the scale anchor to the model reduced the effect of language (from d = .63 to d = .27). A bootstrap analysis with 10,000 bootstrap samples (Preacher and Hayes 2004) yielded a mean indirect effect of –.09. The 95% confidence interval ([–.19, –.01]) excludes 0, demonstrating the significance of this partial mediation.

Using both moderation and mediation, Studies 2 and 3 demonstrate the role of the perceived intensity of the scale anchor in ACE. Thus, these studies provide direct evidence that ACE is indeed an anchor contraction effect.

## *STUDY 4*

The previous studies used nonverbal target stimuli to control for the possible influence of target stimulus language. However, assessing whether ACE is contingent on linguistic features of the to-be-rated object is important both substantively and theoretically. Previous research has shown that L2 to-be-rated (i.e., target) stimuli (e.g., advertising copy) are systematically rated as less emotionally intense than L1 to-be-rated stimuli (Puntoni, De Langhe, and Van Osselaer 2009). Thus, whereas L2 target stimuli yield less emotionally intense ratings than L1 target stimuli (because of a decrease in the perceived emotional intensity of the target itself going from L1 to L2), L2 rating scales yield more emotionally intense ratings than L1 rating scales (because of a decrease in the perceived emotional intensity of the scale anchors going from L1 to L2 and because target stimuli are rated relative to anchors). This raises the question whether the two effects are dependent on or independent of each other. If ACE is driven by the perceived emotional intensity of scale anchors, measures of emotionality should respond in opposite directions to the language of to-be-rated stimuli and of anchoring points. Emotionality ratings should be lower when rating L2 advertisements than when rating L1 advertisements but higher when rating the advertisements using L2 rating scales than when rating L1 rating scales. In Study 4, we manipulate both target and rating scale language and predict for emotionality ratings two main effects in opposite directions.

Study 4 also tests the specificity of ACE to emotional items. According to psycholinguistic accounts for the effect of language on emotional intensity, and stressing the language specificity of autobiographical memories (Harris, Gleason, and Aycicegi 2006; Puntoni, De Langhe, and Van Osselaer 2009), the intensified response to stimuli in a person's native language should be restricted to salient experiential domains such as emotions. Thus, ACE should only occur for emotional items and not for nonemotional assessments (e.g., informativeness). We also explored whether ACE occurs for quality judgments by employing the anchoring points "bad" versus "good," which are not unambiguously emotional or nonemotional.

*Method*

*Design*. The experiment used a 3 (appraisal: emotional intensity vs. informativeness vs. quality) × 2 (language of the advertisements: L1 vs. L2) × 2 (language of the rating scales: L1 vs. L2) mixed design. We manipulated the first two factors within subject and the third between subjects. We counterbalanced the order of ad exposure and language sequence by randomly varying across participants the order in which the advertisements appeared in the booklet and the sequence of L1 and L2 advertisements, leading to eight versions of the questionnaire.

*Procedure*. We recruited participants on campus, and they participated in return for a small reward (N = 155; $M_{age}$ = 22.21, SD = 4.19; 78 women ). Participants were given a booklet containing several advertisements, and we asked them to rate each advertisement on a series of nine-point scales (numbered from 1 to 9). The opening advertisement for all participants was for candies (in Dutch), which we included to familiarize participants with the task. For each advertisement, participants rated four items. The first three were the depen-

dent variables. The L2 positive anchors were "emotional," "informative," and "good." The L1 equivalents ("*emotioneel*," "*informatief*," and "*goed*") were highly similar cognates, making this an especially conservative test. The last item was perceived difficulty, and we added this to confirm that understanding of the L2 advertising copy did not play a role. Finally, we asked participants to answer some demographic questions and to guess the purpose of the study in a written essay.

*Stimuli*. We produced six print advertisements to represent a broad spectrum of advertising appeals and sponsoring organizations. The advertisements featured a fictitious brand name as well as verbal and visual information and promoted a variety of products and causes (adhesive bandages, depression help line, mountaineering equipment, perfume, toaster, and vitamin supplement). For example, in the perfume advertisement, the L2 text was "Caution: May increase heart rate and decrease inhibitions" and was accompanied by the photo of a couple. We initially created the advertisements and the booklets in English. A native Dutch speaker translated the materials to Dutch. A second native Dutch speaker assessed the accuracy of the Dutch translation by a comparison with the back-translation of the materials to English. We randomly created the first ad order and generated the second order by inverting the first one. We similarly generated the two language sequences. In summary, participants rated three advertisements with L2 (English) text and three with L1 (Dutch) text. Half of the participants read the instruction and rating scales in L2 and the other half in L1.

### *Results*

An examination of participants' written essays at the end of the study revealed that none of them guessed the purpose of the research. We jointly subjected the ratings of emotional intensity, informativeness, and quality to a repeated measures ANOVA with language of the advertisements as an additional within-subject factor and language of the rating scales as a between-subjects factor. We make two predictions for this model: (1) The two-way interaction between appraisal and language of the rating scales will be significant, such that advertisements will be rated as more emotional when using L2 rating scales than when using L1 rating scales, with no such effect for nonemotional appraisals, and (2) the two-way interaction between appraisal and language of the advertisements will be significant, such that L1 advertisements will be rated as more emotional than L2 advertisements, with no such effect for nonemotional appraisals (see Table 2 and Figure 3).

The results provide support for our predictions by showing significant interactions between appraisal and language of the rating scales ($F(2, 306) = 3.08$, $p < .05$, $\eta_p^2 = .02$) and between appraisal and language of the advertisements ($F(2, 306) = 4.94$, $p < .01$, $\eta_p^2 = .03$).[3] Importantly, these two-way interactions were not qualified by a three-way interaction of appraisal, language of the advertisements, and language of the rating scales ($p > .67$). To explore the nature of the two-

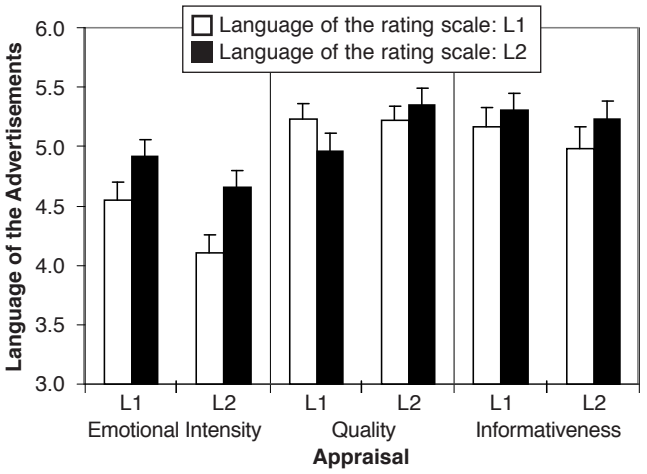
[3]In addition, the main effect of appraisal was significant ($F(2, 306) = 22.97$, $p < .0001$; see Table 2). Driven by the effect of language of the rating scales on emotionality ratings, we also found a marginally significant main effect of the language of the rating scales ($F(1, 153) = 3.05$, $p < .09$). No other effects were significant in this model. Alternative models including the ad order and language sequence counterbalancing factors led to the same results.

Table 2
CELL MEANS (AND STANDARD DEVIATIONS) IN STUDY 4

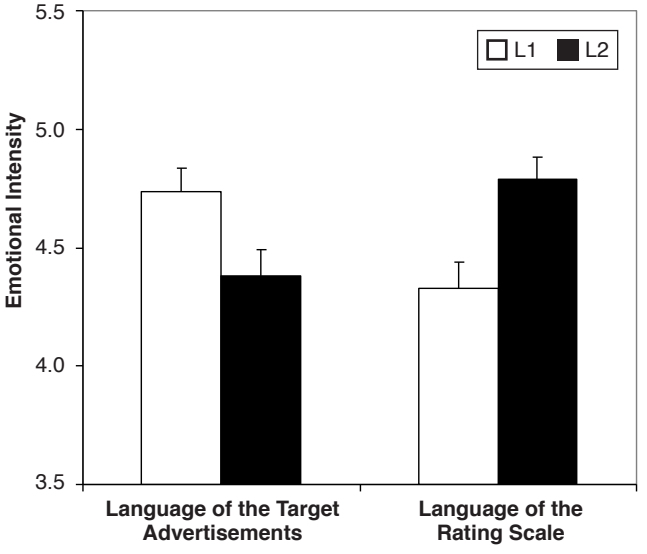
| | *Language of the Rating Scales* | | | |
|---|---|---|---|---|
| | *L1* | | *L2* | |
| | *Language of the Advertisements* | | *Language of the Advertisements* | |
| *Appraisal* | *L1* | *L2* | *L1* | *L2* |
| Emotional intensity | 4.55 (1.32) | 4.10 (1.32) | 4.91 (1.23) | 4.66 (1.27) |
| Informational value | 5.16 (1.40) | 4.99 (1.59) | 5.30 (1.31) | 5.23 (1.32) |
| Quality | 5.23 (1.22) | 5.22 (1.05) | 4.96 (1.34) | 5.35 (1.21) |

Figure 3
APPRAISAL BY LANGUAGE INTERACTIONS FOR LANGUAGE OF THE ADVERTISEMENTS AND LANGUAGE OF THE RATING SCALES IN STUDY 4

*A: Complete Pattern of Means*




*B: Emotionality: Main Effects of Language of the Advertisements and of the Rating Scales*



way interactions, we performed univariate follow-up analyses for each appraisal.

*Emotional intensity*. Replicating ACE, the main effect of the language of the rating scales was significant ($F(1, 153) = 9.79$, $p < .01$, d = .36). Participants rated the advertisements as more emotional when using L2 (M = 4.79) than when using L1 rating scales (M = 4.33). Moreover, the main effect of the language of the advertisements was significant and opposite to that of the language of the rating scales ($F(1, 153) = 5.82$, $p < .05$, $\eta_p^2 = .04$). Participants rated the emotional intensity of the advertisements higher when exposed to L1 advertisements (M = 4.73) than when exposed to L2 advertisements (M = 4.38; see Figure 3, Panel B). The interaction between the two language manipulations was nonsignificant ($p > .51$). In other words, the two language manipulations influenced emotionality ratings independently.

*Alternative appraisals*. The manipulations of the language of the rating scales and the language of the advertisements did not similarly affect the other two appraisals. With one exception, none of the effects of language were significant at the conventional .05 level (for the critical main effect of the language of the rating scales; for informativeness, $p > .30$, and for quality, $p > .67$). We observed a significant interaction between language of the advertisements and language of the rating scales on quality judgments ($F(1, 153) = 4.06$, $p < .05$, $\eta_p^2 = .03$). The features of this interaction indicate higher quality ratings when language of the advertisements and language of the rating scales match (see Table 3 and Figure 3, Panel A). Finally, we found no differences in perceived difficulty of the advertisements as a function of the language manipulations (all *p*s > .44), indicating again that L2 proficiency did not exert a major influence.

### *Discussion*

In this study, participants rated L1 and L2 advertisements using either L1 or L2 rating scales. We predicted opposite effects of ad and scale anchor language on assessments of emotional intensity. As expected, we found higher emotional intensity ratings when the scale was labeled using words in L2 than when using words in L1. Moreover, we observed higher emotionality ratings when the advertisements were in L1 than L2. The latter finding extends the effect of language that Puntoni, De Langhe, and Van Osselaer (2009) observe for slogans and single words to print advertisements. The main effect of language of the rating scales was not significant for the informativeness measure, nor was there a significant effect of language on a rating scale assessing quality using "good" as a scale anchor, which has affective antecedents but does not probe emotions per se. These results might indicate that ACE is limited to words that directly refer to emotions or emotionality.

## *STUDY 5*

Marketing researchers are often interested in probing consumers' likes and dislikes of market offerings. Product evaluation measures vary widely in the extent to which they tap into emotional concepts (Voss, Spangenberg, and Grohmann 2003). For example, a commonly used measure is to ask consumers about the extent to which they love or hate a product (e.g., "I love it" vs. "I hate it"). Other commonly used product evaluation items are much less directly related to emotions or emotionality (e.g., "well made" vs. "poorly made"). In addition, whereas in previous studies we used unipolar emotional scales, bipolar scales (as in the previous example) are also common. For bipolar scales, ACE predicts more intense positive and negative ratings with L2 than with L1 rating scales, as measured by the deviation from the scale midpoint (see Figure 1, Panel B).

Table 3
CELL MEANS (AND STANDARD DEVIATIONS) IN STUDIES 8 AND 9

| *Study (Factor)* | *Levels* | *Language of the Rating Scales* | |
|---|---|---|---|
| | | *L1* | *L2* |
| 8 (Emoticons) | Absent | 2.04 (.75) | 2.52 (.68) |
| | Present | 2.22 (.72) | 2.21 (.64) |
| 9 (Colors) | Absent | 3.23 (.70) | 3.73 (.69) |
| | Present | 3.46 (.76) | 3.46 (.67) |

We recruited 74 college students in a university cafeteria, and they participated in the study in return for a small reward ($M_{age} = 22.52$, SD = 3.90; 33 women). We presented participants with a picture of an armchair and asked them to evaluate the product using an emotional rating scale ranging from "I hate it" (–4) to "I love it" (+4) and another largely nonemotional rating scale ranging from "poorly made" (–4) to "well made" (+4), in English (L2) or in Dutch (L1). Thus, we manipulated the product evaluation measure within subject and language of the rating scales between subjects. We also counterbalanced the order of the product evaluation measure across participants. Consistent with the emotion specificity of ACE found in Study 4, we expect ACE for the former type of measure but not the latter.

We conducted the analysis on the absolute value of the responses, which reflects their extremeness. We ran a repeated measures ANOVA with the emotional versus nonemotional product evaluation measure as a within-subject factor and language of the rating scales and item order as between-subjects factors. As expected, we found a significant two-way interaction between appraisal and language ($F(1, 70) = 6.23$, $p = .01$, $\eta_p^2 = .08$). Participants reported more intense scores for the emotional measure using L2 (M = 2.03, SD = 1.03) than for those using L1 rating scales (M = 1.39, SD = .96; $F(1, 70) = 7.48$, $p < .01$, d = .64), whereas there were no significant differences for the nonemotional measure ($M_{L2} = 1.68$, $SD_{L2} = 1.12$, and $M_{L1} = 1.75$, $SD_{L1} = 1.10$; $p > .79$). The three-way interaction with item order did not qualify this interaction ($p > .60$; the main effects of language and product evaluation measure were also nonsignificant [*p*s > .15]).

This study replicates ACE for commonly used bipolar product evaluation measures using emotional labels ("I love it" vs. "I hate it"). The results confirm that ACE does not extend to product evaluation questions based on "colder" assessments. Finally, we found evidence of ACE only for the emotional item, which argues against a comprehension account (see also Study 4).

## *STUDY 6*

From a substantive point of view, it is important to explore whether ACE occurs across a variety of response

scale formats. Studies 1–4 used unipolar response scales, and Study 5 used a bipolar response scale. However, in all these studies, only the end points of the response scales carried a verbal label, leaving the interpretation of the intermediate response options ambiguous. This raises the question whether ACE also occurs when all individual rating scale points carry verbal labels. To test this, we asked 66 participants ($M_{age}$ = 20.00, SD = 1.65; 25 women) to indicate the extent to which six images from real advertisements made them feel sad or happy. Participants were undergraduate students who took part in the study in exchange for course credit. We used a 2 (language of the rating scale: L1 vs. L2) × 2 (verbal labels: end points vs. all points) between-subjects design with a bipolar rating scale ranging from "very sad" to "very happy" (numbered from –3 to +3). In the only-end-points-labeled condition, these were the only verbal labels used. In the all-points-labeled condition, all response options carried a verbal label ("very sad," "sad," "a little bit sad," "neither sad nor happy," "a little bit happy," "happy," and "very happy"). An ANOVA on the average of the absolute values of the six ratings with language of the rating scale and verbal labels as between-subjects factors yielded a significant main effect of language of the rating scale ($F(1, 62) = 5.20$, $p < .05$, d = .57), which was not moderated by the labeling of end points versus all points ($p > .91$). The main effect of verbal labels was also not significant ($p > .57$). Participants deviated more from the midpoint of the scale when using L2 rating scales (M = 1.49, SD = .43) than when using L1 rating scales (M = 1.23, SD = .51), regardless of whether all or none of the intermediate response options were specified.

## *STUDY 7*

Another substantively important issue is whether ACE is obtained when the emotion is not presented in the response scale itself but only in the question preceding the response scale. For example, in the widely used measure for emotional responses developed by Richins (1997), people rate the intensity of their emotional experiences in consumption situations by answering the question, "To what extent did [situation x] make you feel [emotion y]?" on a four-point scale ("not at all," "a little," "moderately," "strongly"). These anchoring points do not feature emotion words, but respondents must impute the emotion label implicitly to answer the question. To explore the generalizability of ACE to such settings, we asked 76 undergraduate students ($M_{age}$ = 19.66, SD = 2.03; 32 women), after completing an experimental session and in return for course credit, to rate to what extent the session made them feel happy or sad using either L1 or L2 scales from 0 ("not at all") to 6 ("very"). Replicating ACE, a repeated measures ANOVA with emotion (happy vs. sad) as repeated measures and language of the rating scale as a between-subjects factor revealed a main effect of language ($F(1, 74) = 27.23$, $p < .0001$, d = 1.20), in the absence of an interaction between language and emotion ($p > .99$). Participants reported higher levels of happiness ($M_{L2}$ = 4.35, $SD_{L2}$ = 1.25, and $M_{L1}$ = 3.51, $SD_{L1}$ = .72) and higher levels of sadness ($M_{L2}$ = 2.43, $SD_{L2}$ = 1.61, and $M_{L1}$ = 1.59, $SD_{L1}$ = .85) on L2 rating scales than on L1 rating scales (the main effect of emotion was also significant; $F(1, 74) = 83.79$, $p < .0001$). Thus, it is sufficient for the emotion to be presented in the question preceding the rating scale for ACE to emerge.

## *STUDY 8*

Having established ACE for a variety of emotions across a variety of response scale formats, an important open question pertains to the interventions that might mitigate or eliminate ACE. In Studies 8 and 9, we explore the effectiveness of corrective techniques.

Besides attending to verbal cues that communicate the meaning of the response options on a rating scale, respondents' interpretation of a scale is also influenced by pictorial cues (e.g., Tourangeau, Couper, and Conrad 2007). The previous studies established that ACE stems from the greater perceived emotionality of L1 verbal labels. Therefore, one way to mitigate the effect of language of the rating scales could be to provide respondents with an alternative (i.e., nonverbal) basis for interpreting the intensity of the anchoring points. Specifically, language-free cues that provide additional (diagnostic) information about the emotional intensity of scale end points should reduce or eliminate ACE.

In this study, we used emoticons as nonverbal cues. Emoticons are glyphs representing stylized facial expressions used to indicate emotions. In particular, we tested whether adding emoticons to scales measuring specific emotions (happiness and sadness) is sufficient to yield equal intensity ratings on L1 and L2 rating scales. We selected emoticons for three reasons. First, facial expressions are critical tools for conveying emotions. Representations of facial expressions, even if highly stylized, are thus likely to be powerful cues for emotional intensity (e.g., see Walther and D'Addario 2001). Second, emoticons are often used in online environments (Derks, Bos, and Von Grumbkow 2008), in which ACE is especially likely to occur. Third, from an implementational perspective, emoticons for a wide range of specific emotions are readily available for use.

One hundred thirty-two undergraduate students participated in the study in exchange for course credit ($M_{age}$ = 19.87, SD = 2.07; 68 women). Participants watched the short animated movie from Study 2 and rated the extent to which sadness and happiness were portrayed in the movie. Participants provided ratings on two five-point unipolar scales ranging from "no sadness/happiness at all" to "very intense sadness/happiness" (numbered from 0 to 4). We randomly assigned participants to the cells of a 2 (language of the rating scales: L1 vs. L2) × 2 (emoticons: present vs. absent) × 2 (emotion: sadness vs. happiness) mixed design, in which we manipulated language of the rating scales and emoticons between subjects and emotion was a repeated factor. In the emoticons-present condition, we added happy and sad emoticons to the end points of the scales, with a gradual increase in emotional intensity (see Figure 4, Panel A).

A repeated measures ANOVA with language of the rating scales and emoticons as between subjects factors and emotion as a within subject factor (see Table 3) revealed the predicted two-way interaction between language of the rating scales and emoticons ($F(1, 128) = 4.12$, $p < .05$, $\eta_p^2 = .03$).[4] Replicating ACE, when no pictorial cues were present, par-

[4]The three-way interaction between language of the rating scales and emoticons was nonsignificant ($p > .32$). Besides a significant main effect of emotion ($F(1, 128) = 64.57$, $p < .0001$) and a marginally significant main effect of language of the rating scales ($F(1, 128) = 3.70$, $p = .06$), no other effect was significant ($ps > .14$; see Table 3).

Figure 4
CORRECTIVE TECHNIQUES USED IN STUDIES 8 AND 9

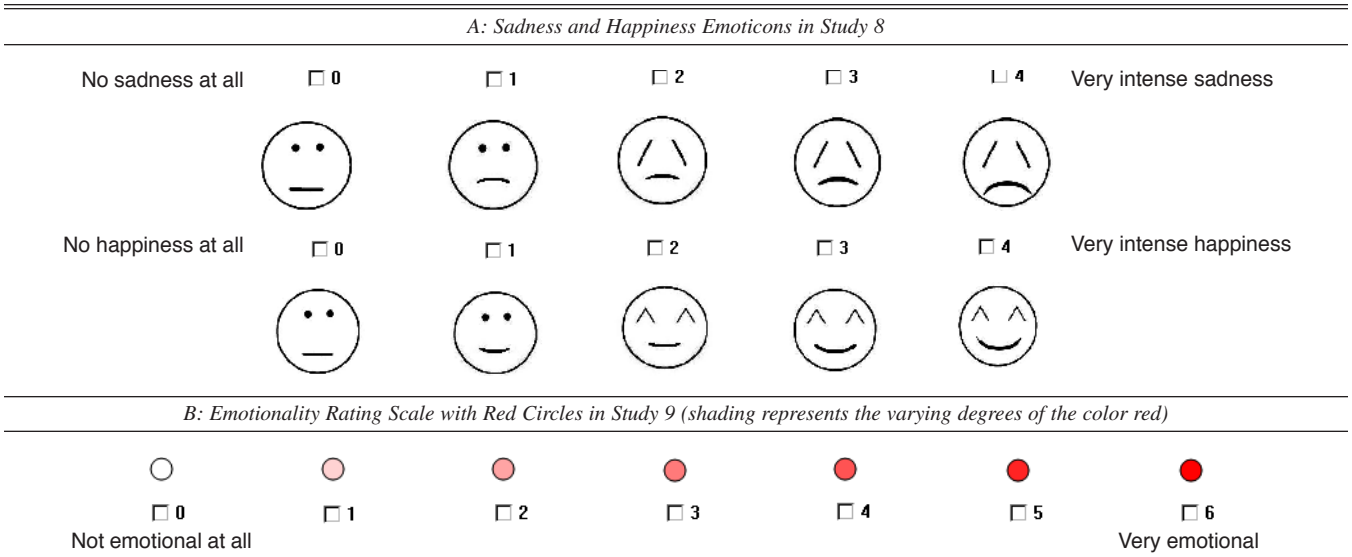
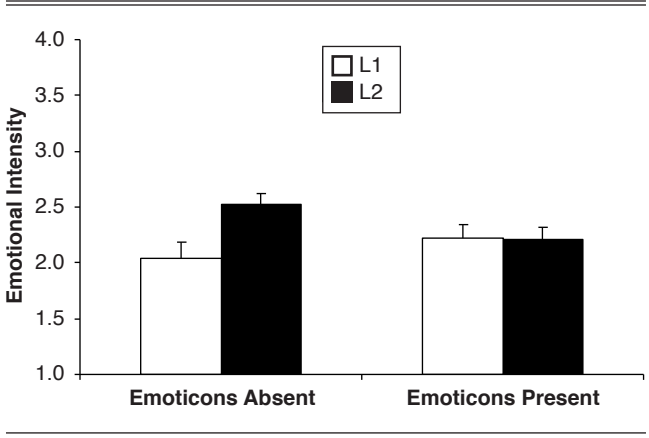



Figure 5
EMOTICONS BY LANGUAGE INTERACTION IN STUDY 8




ticipants provided higher ratings of emotional intensity when rating the items using L2 anchoring points (M = 2.52) than when using L1 anchoring points (M = 2.04; F(1, 128) = 7.59, $p < .01$, d = .70). When we added emoticons ($p > .94$; see Figure 5), we observed no differences between L2 rating scales (M = 2.21) and L1 rating scales (M = 2.22). This study provides evidence that the concomitant presence of nonverbal cues can eliminate ACE. When we added emoticons to the scale, the effect disappeared.

### *STUDY 9*

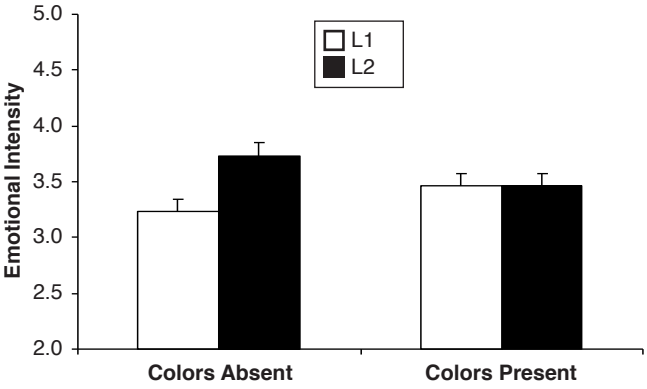
Despite the attractive features of emoticons explored in Study 8, this corrective technique is not applicable to every instance in which ACE could play a role. First, not every emotion can be easily portrayed with emoticons. Second, general indexes of emotional intensity (e.g., the one we used in Study 4) are difficult to represent using facial expressions.

Facial expressions are not the only possible vehicle for emotional information. A promising source of emotional information is color. Colors are strongly associated with emotions (Valdez and Mehrabian 1994), as indicated by the common association of, for example, red with love and anger and blue with depression. In this study, we test the effectiveness of another corrective technique based on the use of colors as cues for emotional intensity. Finding that simply adding color cues with increasing intensity to the rating scales is sufficient to eliminate ACE would be of particular interest from a practical point of view.

We asked 162 undergraduate students ($M_{age}$ = 19.94, SD = 1.49; 47 women), who participated in the study in return for course credit, to evaluate the emotionality of images used in actual skin-care advertisements. We sequentially presented all participants with ten images from which all textual elements had been removed and asked them to rate each image on a scale ranging from "not emotional at all" to "very emotional" (numbered from 0 to 6). We randomly assigned participants to the cells of a 2 (language of rating scale: L1 vs. L2) × 2 (colors: present vs. absent) between-subjects design. For half of the respondents, small red circles of increasing intensity accompanied the verbal scale points (either in L1 or L2; see Figure 4, Panel B). With a degree of simplification, the level of arousal conveyed by a color is a function of its intensity (Valdez and Mehrabian 1994), and the color red is often associated with excitement, stimulation, and arousal (Wexner 1954).

After we averaged across the ten advertisements, an ANOVA with language of the rating scales and color as between-subjects factors (see Table 3 and Figure 6) yielded the expected two-way interaction (F(1, 158) = 4.84, $p < .05$, d = .69). In the absence of color, we replicated ACE. Participants reported higher emotionality ratings with L2 items (M = 3.73) than with L1 items (M = 3.23; F(1, 158) = 9.36, $p < .01$, d = .71). However, when we added color to the scale points, the effect of language of the rating scales was nonsignificant (for both L1 and L2 conditions, M = 3.46, $p > .96$). This study demonstrates that nonverbal cues as simple as colors can eliminate ACE. Together, Studies 8 and 9

Figure 6
COLOR BY LANGUAGE INTERACTION IN STUDY 9




document the effectiveness of easily implementable corrective techniques applicable to virtually all situations in which ACE may be a concern for marketers.

## GENERAL DISCUSSION

One of the most remarkable trends of our time is the increasing globalization of a wide range of economic and social phenomena. Globalization raises new, important questions for marketing researchers. As a consequence, in recent years marketing scholars have begun to explore areas such as information processing in bilingual settings (Luna and Peracchio 2001; Luna, Ringberg, and Peracchio 2008; Noriega and Blair 2008; Puntoni, De Langhe, and Van Osselaer 2009), cross-linguistic issues in marketing communications (Tavassoli and Lee 2003), cross-national logo evaluation (Van der Lans et al. 2009), global consumer culture (Alden, Steenkamp, and Batra 2006), and cross-national invariance of marketing instruments (De Jong et al. 2008). We contribute to this growing literature by exploring the effect of using a nonnative language in marketing research instruments.

In a series of studies, we provide converging evidence for the prediction that bilingual respondents tend to report more intense emotional experiences when using L2 than when using L1 anchoring points (for a summary of the studies, see Table 4). We termed this phenomenon the "anchor contraction effect," or ACE. Studies 1–7 establish ACE by controlling for several different factors. Studies 8 and 9 test two easily implementable corrective techniques. Across the series of studies, Cohen's d for ACE ranged between .36 and 1.20. Therefore, ACE can be characterized as a medium to large effect (Cohen 1992). The studies provide strong evidence of external validity. In Study 1, we observed ACE using a sample of trilingual speakers of Dutch, French, and English, with Dutch and French as target languages. In the remaining studies, we tested ACE using Dutch and English as L1 and L2, respectively. The use of English, the lingua franca of our time (Crystal 1997), as L2 ensures external validity. The use of Dutch, which is relatively close to English (Finegan 1987), as L1 provides a conservative test that enhances the internal validity of the findings. We probed the effect in a variety of settings—movie interpretation (Studies 2 and 8), product evaluation (Study 5), print ad evaluation (Studies 3, 4, 6, and 9), evaluation of an experimental session (Study 7), and taste test (Study 1). We also used several dependent variables—unipolar scales for a range of specific emotions (Studies 1–3 and 7–8), bipolar scales for specific emotions (Studies 5 and 6), and general indexes of emotional intensity (Studies 1, 4, and 9). In addition, ACE was independent of linguistic properties of the to-be-rated stimulus (Study 4). We also obtained ACE when all points of the scale were labeled (Study 6) and when the emotion words had to be imputed implicitly (Study 7).

### *Managerial Implications*

In recent years, the amount of marketing information collected from nonnative speakers has greatly increased. As an example, respondents who are not native speakers of the language in which the questions are formulated routinely provide online customer ratings for a wide variety of products and services. For example, Amazon.com, a global retailer with sales in more than 100 countries, asks customers to rate products using the emotion labels "I love it" and "I hate it." Similarly, Barnesandnoble.com allows visitors from anywhere in the world to rate CDs using the anchoring point "emotional."

What steps should marketers take to control for ACE? The most appropriate solution is to make sure that all respondents answer items in their native language (Kotabe and Helsen 2004). However, providing L1 scales to all respondents can sometimes be too costly or impractical. It is also impossible when the number of native languages in the final sample cannot be predicted beforehand or when respondents from a large number of countries submit ratings (e.g., when polling the inhabitants of multicultural cities such as Chicago, London, and Rotterdam, when a global audience answers questions online).

When the translation approach is not feasible, ACE can be accounted for a priori with corrective techniques. We document the effectiveness of two simple corrective techniques based on the concomitant presentation of verbal and nonverbal cues: emoticons (Study 8) and colors (Study 9). Emoticons can be used when measuring specific emotions, in particular, basic emotions that can be easily portrayed with stylized facial expressions. Emoticons are also especially appropriate in online settings and whenever poor comprehension is a potential concern—such as in the case of children, low levels of L2 proficiency, or low literacy (Kotabe and Helsen 2004). The ease of interpreting emoticons indicates that they may be particularly useful in addressing ACE in emerging markets. Colors are especially suitable in the case of abstract or complex emotional concepts (e.g., "emotional," "pity"), but they may be vulnerable to cross-cultural differences in interpretation (e.g., Roberson, Davies, and Davidoff 2000). Additional research is needed to explore these techniques (1) in other linguistic and cultural contexts; (2) across different formats (e.g., visually marking all points vs. only end points), colors, or emoticons; and (3) across people (e.g., differences in reliance on visual vs. textual information).

If researchers did not implement any of the measures to avoid ACE reviewed in this article at the time of their data collection, they can adopt an a posteriori approach and use information about respondents' L1 as a control variable (e.g., add a dummy variable in regression models). The main drawback of accounting for ACE statistically is that

Table 4
SUMMARY OF STUDIES

| *Study (ACE's d)* | *Task and Target* | *Type of Scale Anchor* | *Type of Emotional Scale Anchor* | *L1* | *L2* | *Scale Characteristics* | | | *Corrective Technique* | *Key Results* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | *Polarity* | *Verbal Labels* | *Target Emotion* | | |
| 1 (.37) | Chocolate taste test | Emotional | Positive | Dutch and French | Dutch and French | Unipolar | Only end points | Only in scale | No | L2 > L1 |
| 2 (.68) | Movie interpretation | Emotional | Positive and negative | Dutch | English | Unipolar | Only end points | In question and scale | No | L2 > L1 |
| 3 (.63) | Ad evaluation | Emotional | Positive | Dutch | English | Unipolar | Only end points | Only in scale | No | L2 > L1 |
| 4 (.36) | Ad evaluation | Emotional and cognitive | Overall emotional intensity | Dutch | English | Unipolar | Only end points | Only in scale | No | Emotional: L2 > L1<br>Cognitive: L2 = L1 |
| 5 (.64) | Product evaluation | Emotional and cognitive | Positive and negative | Dutch | English | Bipolar | Only end points | Only in scale | No | Emotional: \|L2\| > \|L1\|<br>Cognitive: \|L2\| = \|L1\| |
| 6 (.57) | Ad evaluation | Emotional | Positive and negative | Dutch | English | Bipolar | All points | Only in scale | No | \|L2\| > \|L1\| |
| 7 (1.20) | Task evaluation | Emotional | Positive and negative | Dutch | English | Unipolar | Only end points | Only in question | No | L2 > L1 |
| 8 (.70) | Movie interpretation | Emotional | Positive and negative | Dutch | English | Unipolar | Only end points | In question and scale | Emoticons | Emoticons: L2 = L1<br>Control: L2 > L1 |
| 9 (.71) | Ad evaluation | Emotional | Overall emotional intensity | Dutch | English | Unipolar | Only end points | In question and scale | Colors | Colors: L2 = L1<br>Control: L2 > L1 |

the magnitude of ACE is assumed to be the same for all respondents within a language group. This is unlikely to be the case. For example, ACE may depend on the L2 proficiency of the respondent or on the intensity and frequency of prior L2 experiences (Harris, Gleason, and Aycicegi 2006; Puntoni, De Langhe, and Van Osselaer 2009). In addition, (1) this technique imposes additional and often likely impractical burdens on data interpretation for firms, and (2) it may be difficult or impossible to obtain data about respondents' native language.

*Theoretical Implications*

This article contributes to the growing body of work on bilingualism in marketing and consumer research (e.g., Luna, Ringberg, and Peracchio 2008; Noriega and Blair 2008) by highlighting the importance of considering bilingualism in the context of international marketing research. It also contributes to recent research on the emotions of bilinguals (Harris, Gleason, and Aycicegi 2006; Pavlenko 2005; Puntoni, De Langhe, and Van Osselaer 2009) by uncovering a novel consequence of the influence of language on the emotionality of textual information. In particular, our studies provide strong support for the notion that the effect of language of the rating scales on emotionality ratings is driven by a contraction of the scale range at the emotional scale ends of L2 items. Our studies provide process evidence in support of this account (Studies 2 and 3) and rule out alternative explanations, such as language stereotypes (Study 1) and general response tendencies (Studies 4 and 5). In addition, it is worth noting that alternative accounts based on translation issues (i.e., a lack of equivalence between the anchoring points used in the L1 and L2 language conditions) cannot explain results from a balanced bilingual design (Study 1) or from studies using proficient L2 speakers and simple (i.e., easy) words as anchoring points—especially in the case of virtually identical cognates (Studies 1, 2, 4, and 9). Furthermore, the findings for cognates, the interaction effects in Studies 4 and 5, and the perceived difficulty findings in Study 4 rule out an explanation in terms of lack of comprehension of L2 labels. Finally, our results cannot be explained in terms of code switching, or switching between languages, which has been shown to affect the responses of bilinguals (Costa, Santesteban, and Ivanova 2006). In Studies 1, 3, and 4, there was equal code switching for all participants. Moreover, in Study 4, the interaction between the language of the to-be-rated stimuli and of the anchoring points was not significant. In the remaining studies, there was no code switching, because all materials were either in L1 or in L2.

The current studies add to the literature on response styles (e.g., see Baumgartner and Steenkamp 2001; De Jong et al. 2008) by highlighting language (L1 vs. L2) as a determinant of stylistic factors, as well as the content domain (emotional vs. nonemotional) in which this effect occurs. A setting in which these considerations are particularly relevant is cross-cultural research. A standard way to assess cultural influences is to conduct quasi-experiments and compare the answers of respondents with different cultural backgrounds. It is not uncommon for researchers in this area to administer materials in the same language to all participants. In these situations, our studies highlight a threat to the interpretability of data. In addition, information about the language used in the materials is often underreported, making it impossible to assess whether ACE may have played a role in the findings. Thus, we advise researchers interested in measuring emotional constructs across language groups to (1) translate the stimuli into the respondents' native language or use one of the proposed corrective techniques and (2) report the language of the materials.

More research is needed to further explore the boundaries of ACE. In particular, what is an emotional anchor? There is a long-standing debate in the literature about what an emotion is (Frijda 2000). As a result, inventories of emotions tend to differ in both length and content. For a list of consumption-related emotions, we refer to Richins (1997). Clore, Ortony, and Foss (1987) provide a larger emotion lexicon. One reason for the lack of agreement on the definition of an emotion is that many words are considered emotions in some contexts but not in others (Clore, Ortony, and Foss 1987). For example, in Study 4, we found that judgments of an advertisement on a "bad" to "good" scale were not affected by ACE. However, it is possible that if we were to ask respondents how they feel about an advertisement instead of judging the advertisement per se, ACE might arise with the same scale anchors. Similarly, ACE may emerge for "satisfied" in some contexts but not in others, depending on whether the question probes respondents' emotional versus cognitive processes (e.g., "being satisfied" vs. "feeling satisfied"; Clore, Ortony, and Foss 1987).

Another area for further research is the possible individual-level moderators of ACE. For example, greater L2 proficiency has been shown to reduce the magnitude of language effects on emotional responses (Harris, Gleason, and Aycicegi 2006). Thus, effect sizes for ACE may decrease as L2 proficiency approaches that of L1. Finally, we show that ACE also occurs when emotion words are only implicitly featured in an anchoring point, but further research should explore the relevance of ACE in other common response formats, such as Likert scales.

There are currently more L2 than L1 speakers of English, and the number of nonnative English speakers will grow at a rapid pace over the coming decades (Crystal 1997). As information technology enables more people to interact, the amount of data collected from people who are not native speakers of the language of survey questions can only increase. Thus, awareness of ACE and of its remedies is important today and may be critical tomorrow.

## REFERENCES

Alden, Dana L., Jan-Benedict E.M. Steenkamp, and Rajeev Batra (2006), "Consumer Attitudes Toward Marketplace Globalization: Structure, Antecedents, and Consequences," *International Journal of Research in Marketing*, 23 (3), 227–39.

Altarriba, Jeanette (2003), "Does Carino Equal 'Liking'? A Theoretical Approach to Conceptual Nonequivalence Between Languages," *International Journal of Bilingualism*, 7 (3), 305–322.

Anderson, Doug (2009), "Below the Topline: The United States in 2020: A Very Different Place," *SVP Report*, The Nielsen Company, (accessed December 2, 2010), [available at http://en-us.nielsen.com/main/insights/consumer_insight/July_2009/the_united_states].

Baron, Reuben M. and David A. Kenny (1986), "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology,* 51 (6), 1173–82.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (2001), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38 (May), 143–56.

Bruner, Gordon, II, Karen E. James, and Paul J. Hensel, eds. (2001), *Marketing Scales Handbook: A Compilation of Multi-Item Measures*, Vol. 3. Chicago: American Marketing Association.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345–54.

Clore, Gerald L., Andrew Ortony, and Mark A. Foss (1987), "The Psychological Foundations of the Affective Lexicon," *Journal of Personality and Social Psychology*, 53 (4), 751–66.

Cohen, Jacob (1992), "A Power Primer," *Psychological Bulletin*, 112 (1), 155–59.

Costa, Albert, Alfonso Caramazza, and Nuria Sebastián-Gallés (2000), "The Cognate Facilitation Effect: Implications for Models of Lexical Access," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (5), 1283–96.

———, Mikel Santesteban, and Iva Ivanova (2006), "How Do Highly Proficient Bilinguals Control Their Lexicalization Process? Inhibitory and Language-Specific Selection Mechanisms Are Both Functional," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32 (5), 1057–1074.

Crystal, David (1997), *English as a Global Language*. Cambridge, UK: Cambridge University Press.

Cuddy, Amy J.C., Susan T. Fiske, Virginia S.Y. Kwan, Peter Glick, Stephanie Demoulin, Jacques-Philippe Leyens, et al. (2009), "Stereotype Content Model Across Cultures: Towards Universal Similarities and Some Differences," *British Journal of Social Psychology*, 48 (1), 1–33.

De Jong, Martijn G., Jan-Benedict E.M. Steenkamp, Jean-Paul Fox, and Hans Baumgartner (2008), "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation," *Journal of Marketing Research*, 45 (February), 104–115.

Derks, Daantje, Arjan E.R. Bos, and Jasper von Grumbkow (2008), "Emoticons and Online Message Interpretation," *Social Science Computer Review*, 26 (3), 379–88.

Dicken, Peter (2007), *Global Shift: Mapping the Changing Contours of the World Economy*. London: Sage Publications.

Finegan, Edward (1987), "English," in *The World's Major Languages*, Bernard Comrie, ed. Beckenham, UK: Croom Helm, 77–109.

Frijda, Nico H. (2000), "The Psychologists' Point of View," in *Handbook of Emotions*, Michael Lewis and Jeannette M. Haviland-Jones, eds. New York: Guilford Press, 59–74.

Greenleaf, Eric A. (1992), "Measuring Extreme Response Style," *Public Opinion Quarterly*, 56 (Fall), 328–51.

Harris, Catherine L., Ayse Aycicegi, and Jean Berko Gleason (2003), "Taboo Words and Reprimands Elicit Greater Autonomic Reactivity in a First Language than in a Second Language," *Applied Psycholinguistics*, 24 (4), 561–79.

———, Jean Berko Gleason, and Ayse Aycicegi (2006), "When Is a First Language More Emotional? Psychophysiological Evidence from Bilingual Speakers," in *Bilingual Minds: Emotional Experience, Expression, and Representation*, Aneta Pavlenko, ed. Clevedon, UK: Multilingual Matters, 257–311.

Kotabe, Masaaki and Kristiaan Helsen (2004), *Global Marketing Management*, 3d ed. Hoboken, NJ: John Wiley and Sons.

Leclerc, France, Bernd H. Schmitt, and Laurette Dube (1994), "Foreign Branding and Its Effects on Product Perceptions and Attitudes," *Journal of Marketing Research*, 31 (May), 263–70.

Luna, David and Laura A. Peracchio (2001), "Moderators of Language Effects in Advertising to Bilinguals: A Psycholinguistic Approach," *Journal of Consumer Research*, 28 (2), 284–95.

———, Torsten Ringberg, and Laura A. Peracchio (2008), "One Individual, Two Identities: Frame Switching Among Biculturals," *Journal of Consumer Research*, 35 (2), 279–93.

Marian, Viorica and Margarita Kaushanskaya (2004), "Self-Construal and Emotion in Bicultural Bilinguals," *Journal of Memory and Language*, 51 (2), 190–201.

——— and Ulric Neisser (2000), "Language-Dependent Recall of Autobiographical Memories," *Journal of Experimental Psychology: General*, 129 (3), 361–68.

Marschan-Piekkari, Rebecca, Denice Welch, and Lawrence Welch (1999), "Adopting a Common Corporate Language: IHRM Implications," *International Journal of Human Resource Management*, 10 (3), 377–90.

Meyers-Levy, Joan, Rui Zhu, and Lan Jiang (2010), "Context Effects from Bodily Sensations: Examining Bodily Sensations Induced by Flooring and the Moderating Role of Product Viewing Distance," *Journal of Consumer Research*, 37 (June), 1–14.

Noriega, Jaime and Edward Blair (2008), "Advertising to Bilinguals: Does the Language of Advertising Influence the Nature of Thoughts?" *Journal of Marketing*, 72 (September), 69–83.

Ostrom, Thomas M. (1966), "Perspective as an Intervening Construct in the Judgment of Attitude Statements," *Journal of Personality and Social Psychology*, 3 (2), 135–44.

Pavlenko, Aneta (2005), *Emotions and Multilingualism*. Cambridge, UK: Cambridge University Press.

Pew Research Center for the People and the Press (2001), "October 2001 News Interest Update," (accessed January 11, 2010), [available at http://people-press.org].

Preacher, Kristopher J. and Andrew F. Hayes (2004), "SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models," *Behavior Research Methods Instruments & Computers*, 36 (4), 717–31.

Puntoni, Stefano, Bart de Langhe, and Stijn M.J. van Osselaer (2009), "Bilingualism and the Emotional Intensity of Advertising Language," *Journal of Consumer Research*, 35 (6), 1012–1025.

Richard, Marie-Odile and Roy Toffoli (2009), "Language Influence in Response to Questionnaires by Bilingual Respondents: A Test of the Whorfian Hypothesis," *Journal of Business Research*, 62 (10), 987–94.

Richins, Marsha L. (1997), "Measuring Emotions in the Consumption Experience," *Journal of Consumer Research*, 24 (September), 127–46.

Roberson, Debi, Ian Davies, and Jules Davidoff (2000), "Color Categories Are Not Universal: Replications and New Evidence from a Stone-Age Culture," *Journal of Experimental Psychology: General*, 129 (3), 369–98.

Schwarz, Norbert, Bärbel Knauper, Hans J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clark (1991), "Rating-Scales: Numeric Values May Change the Meaning of Scale Labels," *Public Opinion Quarterly*, 55 (4), 570–82.

Sunderman, Gretchen and Judith F. Kroll (2006), "First Language Activation During Second Language Lexical Processing: An Investigation of Lexical Form, Meaning, and Grammatical Class," *Studies in Second Language Acquisition*, 28 (3), 387–422.

Tavassoli, Nader T. and Yih Hwai Lee (2003), "The Differential Interaction of Auditory and Visual Advertising Elements with Chinese and English," *Journal of Marketing Research*, 40 (November), 468–80.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad (2007), "Color, Labels, and Interpretive Heuristics for Response Scales," *Public Opinion Quarterly*, 71 (1), 91–112.

Upshaw, Harry S. (1965), "The Effect of Variable Perspectives on Judgments of Opinion Statements for Thurstone Scales: Equal-Appearing Intervals," *Journal of Personality and Social Psychology*, 2 (1), 60–69.

Valdez, Patricia and Albert Mehrabian (1994), "Effects of Color on Emotions," *Journal of Experimental Psychology: General*, 123 (4), 394–409.

Van der Lans, Ralf, Joseph A. Cote, Catherine A. Cole, Siew Meng Leong, Ale Smidts, Pamela W. Henderson, et al. (2009), "Cross-National Logo Evaluation Analysis: An Individual-Level Analysis," *Marketing Science*, 28 (September/October), 968–85.

Van Rosmalen, Joost, Hester van Herk, and Patrick J.F. Groenen (2010), "Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model," *Journal of Marketing Research*, 47 (February), 157–72.

Voss, Kevin E., Eric R. Spangenberg, and Bianca Grohmann (2003), "Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude," *Journal of Marketing Research*, 40 (August), 310–20.

Walther, Joseph B. and Kyle P. D'Addario (2001), "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication," *Social Science Computer Review*, 19 (3), 324–47.

Wexner, Louis B. (1954), "The Degree to Which Colors (Hues) Are Associated with Mood-Tones," *Journal of Applied Psychology*, 38 (6), 432–36.

Wildt, Albert R. and Michael B. Mazis (1978), "Determinants of Scale Response: Label Versus Position," *Journal of Marketing Research*, 15 (May), 261–67.