

Journal of Experimental Social Psychology, in press

The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude
valence, extremity, and emotionality

Matthew D. Rocklage^a and Russell H. Fazio^{a*}

^aOhio State University, 1835 Neil Avenue, Columbus, OH 43210, USA

* Corresponding author at: Department of Psychology, Ohio State University, 1835 Neil
Avenue, Columbus, OH 43210, USA. Fax: +1 614 292 5601

E-mail addresses: rocklage.1@osu.edu (M. D. Rocklage), fazio.11@osu.edu (R.H. Fazio)

ABSTRACT

The language – in particular, the adjectives – individuals use can be harnessed to understand the different aspects of their attitudes. The present research introduces a novel approach to measuring attitudes that allows researchers to quantify these aspects. In Study 1, we created a list of 94 evaluative adjectives and asked participant judges to rate the implied valence, extremity, and emotionality of each adjective. This approach allowed us to quantify each adjective along these dimensions and thereby create the Evaluative Lexicon (EL). We validated the EL in Study 2 by experimentally creating attitudes toward novel stimuli in the lab and then measuring them using our tool. In Study 3, we sought to further validate the EL as well as demonstrate its practical and theoretical contributions using a natural-text repository of 5.9 million Amazon.com product reviews. Results from the Amazon.com reviews indicate that individuals use the EL adjectives in ways that further validate their ability and usefulness in measuring valence, extremity, and emotionality even within natural text. These findings, in turn, produced new theoretical contributions regarding the separability of attitude extremity and emotionality as well as their relation to summaries of both univalent and ambivalent evaluations. The findings highlight the importance of attitude emotionality for understanding attitude expressions.

Keywords: attitudes, attitude measurement, affect, cognition, text analysis

The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude
valence, extremity, and emotionality

Words are of immense importance to our understanding of others. They provide a window into people's thoughts and feelings, their intentions and their biases. As such they are of paramount significance when seeking to understand others' perceptions of the world (Holtgraves, 2010). As we think about this world of words, however, it becomes readily apparent that there exists a multitude of words that, at face value, appear to convey similar meanings. For example, we could always just say we "like" or "dislike" an object when we evaluate it. Instead, we turn to words such as "beneficial" or "harmful," or perhaps even "wonderful" or "disgusting." We seem to have an abundance of ways to express the general positivity or negativity we have associated with an object. Why might this be? Why does our language provide us with so many similar words to convey our liking or disliking?

Given this wide range of available language, it appears that our words have more or less subtle distinctions that help us to communicate our internal thoughts and emotions. For this particular paper, we are interested in what these distinctions might tell us about individuals' underlying attitudes. For example, when we describe our smartphone as "wonderful" instead of "helpful," what does that say about the attitudes we hold? One possibility we wish to focus on is that when we use the word "wonderful," we are describing an attitude that is not only more extreme, but also one that has some basis on emotion. The word "helpful," however, is one that may be primarily based on cognition – that is, beliefs about the object and its properties.

This distinction between attitudes based on affect versus those based on cognition has been of great interest to researchers for a number of decades both for its theoretical and practical implications. For example, researchers interested in prejudice measured the cognitive basis of

individuals' attitudes toward minority groups by asking them to list their stereotypes and symbolic beliefs (e.g., values that are facilitated or hindered by the target group; Haddock, Zanna, & Esses, 1993). These researchers then measured individuals' affective basis by having them list the feelings or emotions they experienced when seeing, meeting, or thinking about members of this group. When predicting overall attitudes toward minority groups, they found that both the cognitive and affective bases were significant contributors to individuals' summary evaluations. Additionally, researchers have found that individuals are relatively more persuaded by an argument when that argument's appeals match the basis of individuals' attitudes for that attitude object (Fabrigar & Petty, 1999). When their attitude is based primarily on affect, individuals are relatively more persuaded by arguments that are also based on affect, and vice-versa for more cognitively-based attitudes. Relatedly, from an individual difference perspective, individuals who generally base their attitudes on affect across a range of attitude objects are more persuaded by an affective appeal for a novel product while those who tend to base their attitudes more on cognition are less persuaded by this same appeal (Huskinson & Haddock, 2004). Finally, it has been theorized, and some evidence found, that affectively-based attitudes are more accessible in memory than cognitively-based attitudes (Fazio, 1995; Van den Berg, Manstead, van der Pligt, & Wigboldus, 2006). Attitudes that are more accessible are more likely to direct attention, more stable over time, and more likely to drive subsequent behavior (Fazio, 1995). All of these findings demonstrate the importance of understanding the contribution of affective versus cognitive bases to the attitudes that people develop.

Given this interest in the affective and cognitive bases of attitudes, different approaches also exist to measure these bases. Early research, for example, often utilized study-specific measures that did not easily transfer to different attitude objects. These scales were often tailored

to a single object. For example, when eliciting evaluations of political candidates, researchers asked participants to state, for instance, whether the candidates made them feel “angry” or “hopeful” to measure the affective basis of individuals’ attitudes, and how “honest” and “knowledgeable” they seemed to measure the cognitive basis (e.g., Abelson, Kinder, Peters, & Fiske, 1982; Lavine, Thomsen, Zanna, & Borgida, 1998). As is apparent, these words would not readily apply to a diverse range of attitude objects. The difficulty in applying these idiosyncratic scales to different objects also meant that it would be difficult to compare results across studies. Finally, this diversity of scales also led to idiosyncratic definitions of affectively-based attitudes. While some studies focused on emotion-related terms, others equated affect with general evaluations (e.g., very favorable to very unfavorable; Norman, 1975).

Due to these issues, researchers have sought solutions for measuring attitude bases that can be compared across a wide range of attitude objects. These solutions have tended to fall into two different categories. One approach is more open-ended (e.g., Eagly, Mladinic, & Otto, 1994; Haddock, Zanna, & Esses, 1993; Haddock & Zanna, 1998) and the other is more “closed” (e.g., Crites, Fabrigar, & Petty, 1994). The open-ended approach requires that individuals introspect on each aspect of their attitudes singly and create a list of their beliefs and a separate list of their emotions they had regarding a particular attitude object. After creating these lists, participants would then rate the implied valence associated with each belief and emotion they listed. Researchers could then calculate the average valence associated with each list to predict individuals’ scalar ratings of their overall summary attitudes and, in that way, understand whether attitudes toward the particular object were generally based relatively more on affect or cognition. The more “closed” approach, on the other hand, avoided the step of asking respondents to list relevant beliefs and emotions. Instead, using the Crites et al. measure,

participants responded to a set of three scales: one set focused on measuring the affective basis, one on the cognitive basis, and one measuring individuals' summary evaluation. The affect semantic differential scales required individuals to select a position on a scale that "best described their feelings toward the object" (e.g., love/hateful, delighted/sad, acceptance/disgusted, etc) while the cognitive scales required participants to select a position on the scale that "best described the traits or characteristics of the object" (e.g., useful/useless, wise/foolish, beneficial/harmful, etc). Each set of scales, then, required that individuals assess and report on the different possible bases of their attitudes toward an object.

In the current paper, we considered another way to assess attitudes and their bases in an even richer and more flexible manner and one that would allow researchers to test new hypotheses that lead to advances in social psychological theory. In particular, we became interested in the varying connotations of evaluative adjectives, as exemplified, for instance, by the semantic anchors utilized by Crites et al. (1994). They used words like those we mentioned at the beginning of this paper: "beneficial," "harmful," "disgusted," etc. Researchers' intuitive use of these various words, in and of itself, suggests that the words may differ in the extent to which they imply evaluations based on affect or cognition. Would it be possible to analyze words like these and then use that information to better understand individuals' attitudes toward different objects? Is it possible to quantify adjectives like these and what they imply so that when individuals use them we can obtain an enhanced understanding of their underlying attitudes?

Concentrating on these kinds of words and quantifying them would allow us to create a tool that has a number of methodological benefits. First, if we were to simply have participants select words from a predefined list, we would only have to ask them a single question: "which of the following words best describe your attitude toward this object?" As noted earlier, measures

of attitude basis often require individuals to introspect on their emotions and beliefs and then list them one-by-one (e.g., Eagly et al., 1994; Haddock et al., 1993). This is likely quite difficult for many participants and, indeed, can lead to missing data from individuals who struggle to list any reactions at all (as in Crites et al., 1994; see also Haddock & Zanna, 1998). Instead of responding to direct questions that require them to introspect and focus on either the affective or the cognitive basis of their attitudes singly, participants could simply select those words that best reflected their attitude. Our approach would give individuals the freedom to choose words – ones that imply affect to differential degrees – that seemed to fit their evaluation regardless of the evaluation's affective and/or cognitive basis. In essence, we could utilize individuals' naturalistic use of different evaluative terms and leverage the denotations and connotations of those specific words. Second, we could increase the efficiency of measuring individuals' attitudes. The open-ended measures require a great deal of time and effort for the respondents to introspect, enumerate, and then rate their emotions and beliefs. The Crites et al. (1994) measure requires participants to respond to numerous separate scales for each attitude object (19 in that specific research), some of which focus attention on feelings toward the object in question and some of which focus attention on the traits or characteristics of the object. Our approach would again ask only a single question with a limited number of response options. Furthermore, given that a single word can communicate multiple aspects of an individual's attitude, a minimal number of selections are required that nevertheless have the potential to provide information regarding multiple variables. Third, this approach would also allow us to measure individuals' attitudes across a range of settings. While the tool we create could be used within a laboratory setting by having individuals select those words that best represent their attitude from a list, it is flexible in that it could also be used to analyze pre-existing text or speech databases that are evaluative in

nature (e.g., Amazon.com product reviews, as we demonstrate in Study 3). Expanding the range of domains that researchers could use to measure attitudes and their bases would allow for converging evidence across multiple, diverse samples and enhance the potential for asking new and interesting questions.

Importantly, such an approach could also help to bring attention to an overlooked distinction in the attitudes literature, that between extremity and emotionality. Indeed, to our knowledge, there has not been an attempt to distinguish these two constructs empirically. For instance, past research has suggested that when positivity and negativity conflict, the valence that is based more on emotion tends to be more related to summary evaluations (Lavine et al., 1998). The conclusion that emotionality in-and-of itself is integral to these effects, however, may not be warranted. This research assessed emotionality by whether participants indicated “yes” or “no” if they felt hopeful, proud, angry, afraid, etc in regard to a political figure. Yet, indicating “yes” to an emotion may also indicate greater extremity of the associated valence. To say that a politician makes one feel, for example, “angry” suggests extreme negativity toward the politician. Is it the emotionality or the extremity that matters? That is, did the negativity relate more to the summary evaluation than the conflicting positivity because the negativity was more emotional or because it was more extreme?

Emotional reactions may influence summary judgments more readily due to their extremity. However, there may be other sources of emotionality’s influence as well. Indeed, emotionality’s influence may also stem from its perceived diagnosticity (Fazio, 1995). Because emotion emanates from the person him/herself, it can be perceived by that person as all the more diagnostic of his/her attitude. Contrast this to a more cognitively-based attitude, which relies relatively more on probabilistic knowledge one has acquired about the properties of an object

and, hence, is less related to the person him/herself. The perceived diagnosticity of emotionality, in turn, may influence individuals as they arrive at a summary evaluation of the object.

This is all to simply point out that emotionality and extremity likely covary and yet, theoretically, we can assess the separable influences of the two. If adjectives connote not only valence, but also the emotionality and extremity of that valence, then it should be possible to assess the unique contributions of both emotionality and extremity.

Our goal for the current paper is to validate a general approach to measuring individuals' attitudes through their use of different words and then use this approach to investigate the possibility that there are separable effects of attitude extremity and emotionality. To do so, we looked to enumerate and then quantify a list of words that are often used to describe evaluations of a wide range of objects. We therefore focused on those words that are descriptive by their very nature: adjectives. Adjectives are particularly important when expressing attitudes because they are essentially communicators' attempts to describe an object and their favorability toward it (e.g., as beneficial, terrible, fantastic). Our general approach is to first obtain normative ratings of each adjective's implied valence, extremity, and emotionality, and then use these normative ratings in place of that adjective whenever individuals use it to express their evaluations.

Study 1

In this first study our goal was to create a list of adjectives that could be quantified and then utilized as a tool for measuring attitudes and their basis in subsequent research.

METHOD

Judges

One hundred and forty participant judges from the Mechanical Turk website participated. Judges ranged from 18 to 73 years old ($M = 38$). The data from seven were excluded because

they failed an attention check within the survey. One other was excluded for not being a native English speaker. A total of 132 judges remained for subsequent analyses (70 males; 62 females).

Procedure

Because we are interested in using them as a tool across a wide range of possible attitude objects, the adjectives we identified predominantly originated from a large list of adjectives that had been used in past work in our lab as stimuli for the evaluative priming procedure (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The adjectives were required to fit a number of characteristics for inclusion in the current project. First, they were required to be evaluative in nature and have obvious positive or negative denotations. Adjectives such as “big” or “typical” have no consistent valence associated with them and therefore were not included on our list. Second, we selected adjectives that seemed to represent a wide range of valence, extremity, and emotionality. In regard to emotionality, for example, the words “wise” and “terrific” are both clearly positive, but seem to differ in the extent to which they imply emotionality. Finally, we selected adjectives that were applicable across multiple domains. The word “terrific,” for example, could be used to describe stimuli in general – anything from a person, place, object, idea, or action. This criterion eliminated adjectives that provided very specific denotative meanings – e.g., trait terms such as “honest” which apply solely to people. The end result was a list of 94 adjectives that was reasonably exhaustive with respect to the adjectives having both an evaluative connotation and broad applicability (see the Appendix for all adjectives).

Judges were asked to rate each of the 94 evaluative adjectives either on their implied valence ($N = 68$) or their implied emotionality ($N = 64$). Those rating implied valence were given the instructions:

When we evaluate an object, person, or event, we often use terms such as those listed on the next page. Some evaluative terms imply a negative evaluation and some a positive evaluation. Using the scale on the next page, please rate the evaluation implied by each term.

Next to each of the 94 adjectives, judges were given a 0 (Very negative) to 9 (Very positive) scale to rate each adjective.

Those rating implied emotionality were given the instructions:

Sometimes when we evaluate an object, person, or event, we do so on the basis of an emotional reaction to the object, person, or event. That is, our emotions determine whether we conclude that we like or dislike the object, person, or event. Indeed, some evaluative terms appear to imply that the evaluation was arrived at on the basis of emotion. Using the scale on the next page, please rate the extent to which each term implies an evaluation based on emotion.

Judges were given a 0 (Not at all emotional) to 9 (Very emotional) scale to rate each adjective.

RESULTS

Three variables of interest can be extracted from our measures: valence, emotionality, and extremity. To compute the implied valence and emotionality for each adjective we simply averaged these scores across participants. At the level of participants' individual ratings, extremity was calculated as the absolute value of the valence rating minus the midpoint of the scale ($abs(\text{Valence} - 4.5)$). This deviation from the midpoint was then averaged across participants for each adjective.

The means and standard deviations for each of the three variables are presented for each adjective in the Appendix. The averaged valence ratings for the adjectives ranged from 0.40 to 8.71 ($M = 4.55$; $SD = 3.11$), emotionality ratings ranged from 2.45 to 7.61 ($M = 5.27$; $SD = 1.45$), and extremity ratings ranged from .63 to 4.21 ($M = 3.04$; $SD = .80$). These wide ranges indicate there is substantial variability in the connotations of the various adjectives. The adjectives differ not only in the favorability that they imply, but also the emotionality that they connote.¹

We can also investigate the associations between the variables. As two different groups of participants rated valence and emotionality separately, we computed correlations at the level of the adjectives averaged across the raters. As would be expected from a list of adjectives that are representative of a wide range of valence and emotionality, there was no correlation between these two variables ($r(92) = -.14$, $p = .18$). We can investigate this same relation between valence and extremity. There was also no significant correlation ($r(92) = -.06$, $p = .56$). These analyses demonstrate that the positive and negative adjectives in our list do not differ in the extent to which they imply an attitude based on emotion or the extent to which they imply extremity.

Although the past literature has provided no direct evidence that it is the case, intuitively, emotionality likely covaries with extremity. We can assess the relationship between these two variables for our adjectives. Indeed, there was an association such that adjectives rated as more emotional by the one set of judges were also rated as more extreme by the other set of judges ($r(92) = .63$, $p < .001$). Although there is a moderate association between these two variables, it is readily apparent that they are not the same. For example, although they are approximately

¹ The reader may also be interested in the coherence of judgments between males and females. To assess this we calculated means for the adjectives separately for male and female judges. The correlation between these two sets was extremely high for both emotionality ($r(92) = .92$, $p < .001$) and valence ($r(92) = .99$, $p < .001$), indicating that males and females converged with respect to their ordering of the adjectives on the dimensions of interest.

equivalent in their extremity, the positive adjectives “perfect,” “excellent,” and “magnificent” are increasingly emotional. Likewise, “useless,” “terrible,” and “repulsive” vary considerably in their implied emotionality, despite being rated similarly with respect to their extremity.

DISCUSSION

The results of Study 1 provide a first step toward validating our general approach. In this study we enumerated a broad set of adjectives and then quantified these adjectives along three dimensions using a large sample of participant judges. The Evaluative Lexicon (EL) included information regarding the valence, extremity, and implied emotionality of 94 adjectives.

It is important to point out that the ratings elicited from the separate sets of judges corresponded in important and meaningful ways. For the creation of a tool, it was important that there not be correlations between valence and either extremity or emotionality, which could bias subsequent analyses using the EL. Given the null correlation between valence and these other two variables, we can say that at least at the level of our adjectives there appears to be no inherent bias for positivity to be more extreme or emotional than negativity or vice versa. Also important, however, is that there was a significant correlation between extremity and emotionality. This finding accords with the intuition that stronger emotional reactions should be associated with more extreme evaluations. The correlation is not so strong, however, as to indicate that these two variables represent the same construct. To our knowledge, this is the first evidence showing that attitude emotionality and extremity covary, but that they are separable.

More generally, it is also worth noting that the results of our study indicate that individuals were able to judge the implied valence, extremity, and emotionality of adjectives. Indeed, many of the words that the judges rated as highly emotional were also words that have been used in past research to anchor scales measuring individuals’ affectively-based attitudes

(e.g., “disgusted,” “excited,” “joyful”), while those rated as relatively low in implied emotionality corresponded to past scales measuring individuals’ cognitively-based attitudes (e.g., “useful,” “beneficial,” “unhealthy”). At the very least, then, it appears that when asked to analyze these words, individuals are able to discern their meaning across multiple dimensions and that this understanding matches that of previous researchers. Whether individuals’ actual use of evaluative adjectives provides information about their attitudes is the subject of our subsequent studies in this paper.

Study 2

Although individuals were able to meaningfully order the adjectives based on their valence, extremity, and emotionality when asked to, it is less clear whether these adjectives accurately reflect individuals’ underlying attitudes. Particularly important for the current study was that our general approach could distinguish not only the valence of an individual’s attitude, but also the nature of the information upon which individuals are basing their attitudes. To this end, we looked to further validate our approach by experimentally creating attitudes in the laboratory that varied in known ways on valence as well as the extent to which they were based on cognition versus affect. We would then have participants select the adjectives that best described that newly-created attitude. If the normative values we obtained from judges for each adjective from Study 1 could be used to identify both the valence and the basis of the attitude we created for participants, this would further validate our approach and the EL itself. In addition, we sought to obtain further evidence of the dissociability of emotionality and extremity when predicting these created attitudes.

METHOD

Participants

Sixty-nine undergraduate students participated in partial fulfillment of a psychology course requirement or for payment. One participant was excluded for failing to follow instructions. A total of 68 participants remained for subsequent analyses.

Procedure

Participants began the experiment by reviewing a list of adjectives from the EL, with which they were asked to familiarize themselves for use later in the experiment. A representative subset of about half of our original adjectives was used to shorten the experiment's overall length and simplify the participants' task. This particular subset of adjectives was chosen in such a way so as to ensure coverage of the range of valence and emotionality present in the original EL.² Given that the adjectives we included did not differ from those we did not include, we are, in essence, testing whether the general approach of obtaining normative ratings for valence and emotionality is appropriate and thereby validating the lexicon in its entirety.

After familiarizing themselves with this list, participants were then asked to consider their evaluation of an aquatic animal, the "lemphur," which they would be reading about momentarily (Crites et al., 1994). They were told that after they had finished reading this passage describing the lemphur, they would be asked to choose two to five adjectives that describe their evaluation of the lemphur. They were also told they would be asked to indicate which *one* of those two to five adjectives *best* described their evaluation of the lemphur.

² Comparison of the subset of adjectives used in this study with those not used revealed no differences in valence, extremity, or emotionality ($ps > .31$). Minimum and maximum values were also similar. As with the original list of EL adjectives, the adjectives chosen showed no association between emotionality and valence ($r(41) = -.06, p = .75$) or extremity and valence ($r(41) = -.06, p = .70$) and a comparable correlation between emotionality and extremity ($r(41) = .52, p < .001$). See the Appendix for the specific adjectives used.

Participants were randomly assigned to one of four conditions in a 2 (basis: cognitive or affective) X 2 (valence: negative or positive) factorial design, each with a passage describing the fictitious lemphur. All passages were modified versions of their originals from Crites et al. (1994).³ The two passages that were designed to create cognitively-based attitudes described the lemphur in encyclopedic terms explaining that lemphurs, as one example, tend to deplete fish near coastal communities that are dependent on fishing (negative, cognitive passage) or that lemphurs were a source of material for both clothing products and nutrition (positive, cognitive passage). The two passages that were designed to create affectively-based attitudes were short narratives each describing a swimmer's encounter with a lemphur. In the affective, negative condition, the passage was of a swimmer being attacked, mutilated, and dismembered by a lemphur, whereas the affective, positive condition recounted a swimmer frolicking and soaring through the water with a lemphur.

Following the passage, participants selected the two to five adjectives that described their evaluation of the lemphur and then selected *the* adjective of those that *best* described their evaluation.

RESULTS

Given that emotionality and valence ratings had been obtained in Study 1, each chosen adjective had normative values that could be imputed in its place. For instance, if participants chose the adjective "terrifying" to describe their evaluation of the lemphur, the word "terrifying" could then be looked up in the EL and be represented by two scores: valence (.72) and

³ Each passage was modified in order to remove any use of words that were either included in the original EL or could be considered synonymous. This helped to ensure that individuals were not simply selecting adjectives that appeared in the passages and instead that selection was based on the underlying evaluation we created.

emotionality (7.50). These values could then be analyzed to understand participants' underlying evaluations.

The data were analyzed in two ways. The first way was to simply average across all selected adjectives' implied valence and emotionality to create a mean for each of these dimensions for each participant. The second way was to use just participants' self-selected "best" adjective. Each of these approaches will be discussed in turn.

Average of the selected adjectives. To predict participants' condition, and thus the nature of their attitude toward lemphurs, we utilized discriminant function analysis. Discriminant analysis works by combining predictors into latent variables – called functions – and then using these functions so as to minimize the misclassification of cases into their respective groups or conditions. The better these functions are at predicting each group, the more accurate the resulting classification will be. In this case, we began by predicting which of the four experimental conditions participants had experienced from their averaged valence and emotionality across the two to five adjectives they chose.

The discriminant analysis resulted in two functions, with valence and emotionality having their strongest loadings on separate functions (.98 & 1.00, respectively) and with little loading on the other variable's function (.04 & -.21, respectively). Both the function associated with valence ($\chi^2(6) = 225.20, p < .001$) and that associated with emotionality ($\chi^2(2) = 66.44, p < .001$) contributed significantly to the prediction of condition. Of primary concern, however, was our adjectives' overall effectiveness in discriminating between the conditions. Overall, the two

functions combined were able to classify 88.2% of the participants accurately, representing a 63.2% improvement over chance (25%).⁴

We can also ask the extent to which valence by itself is able to discriminate between passages that are either positive or negative. In other words, looking at our cognitive and affective conditions separately, how well does valence predict participants' original condition? Results indicated that implied valence was an extremely good predictor of whether participants received a positive or negative passage within just the affective passages with 100.0% of participants being correctly classified. Within the cognitive passages 97.1% of participants were correctly classified. Both numbers are much better than chance (50.0%). These extremely high accuracy rates make sense as participants are very unlikely to select positive adjectives to represent negative attitudes and vice-versa.

More interestingly, we can also assess the ability of our adjectives to discriminate between cognitively- and affectively-based attitudes within a given valence. In other words, do the adjectives that participants selected reveal whether they had been exposed to the cognitive or the affective passage? To this end, we examined the positive and negative passages separately and utilized participants' averaged emotionality scores to predict their attitude basis. Results indicated that we were successful in discriminating between the cognitive and affective passages for 93.9% of the participants who had been exposed to a positive passage and for 85.7% of the

⁴ For readers who are interested, we also conducted a 2 (valence) X 2 (attitude basis) analysis of variance (ANOVA) assessing the mean implied valence and emotionality of the adjectives participants selected in each condition. These results conceptually replicated those provided by the discriminant analysis. More positive (negative) adjectives were used to describe individuals' evaluations in the positive (negative) conditions and more emotional (cognitive) adjectives in the emotional (cognitive) conditions. See the Supplementary Materials for more detail.

participants who had been exposed to a negative passage. In both cases, this classification rate is considerably better than what would be expected by chance (50.0%).

Given that one of our aims is to illustrate the distinction between emotionality and extremity, we ran the same model, but this time using implied extremity instead of emotionality. This model resulted in two functions with valence and extremity having their strongest loadings on separate functions (.97 & 1.00, respectively) and with little loading on the other variable's function (-.24 & -.02, respectively). Both the function associated with valence ($\chi^2(6) = 184.43, p < .001$) and extremity ($\chi^2(2) = 24.51, p < .001$) contributed significantly to the prediction of condition. Overall, the two functions combined were able to classify 70.6% of participants correctly. Simply comparing the two discriminant function analyses conducted, extremity's ability to differentiate between the conditions was not as good as had been true of emotionality's (70.6% versus 88.2%).

To directly test the differential ability of extremity and emotionality in predicting individuals' attitudes and to further examine their separability, we then included all three variables – valence, extremity, and emotionality – in the same model. As a first indication of extremity and emotionality's differentiation, the discriminant model resulted in three different functions: one most associated with valence (loading of .97), one most with extremity (.88), and one most with emotionality (.99). While valence was largely unrelated to extremity and emotionality (-.02 & .03, respectively), extremity and emotionality did show some association with each other (.48 & .14 on each other's functions, respectively), as we would expect. Most importantly, the function associated with emotionality remained significant even when including extremity ($\chi^2(2) = 70.18, p < .001$) while the function associated with extremity fell to non-

significance ($\chi^2(1) = 3.41, p > .05$). The overall percentage of participants correctly classified remained at 88.2%, no different than when including just valence and emotionality.

These results indicate that emotionality and extremity are separable and that, in the case of our lemphur passages, emotionality is the better predictor. Having set out to manipulate valence and attitude basis, not necessarily extremity, we would anticipate that these variables would be better predictors than extremity. For example, the adjective “useless” was often used to describe the lemphur that destroys coastal food sources while the word “awful” was often used to describe the lemphur that mutilated the swimmer. While these adjectives are nearly equivalent in extremity (~ 3.75), they differ markedly in their implied emotionality (4.25 versus 6.61, respectively) and therefore emotionality would be best suited in discriminating between the passages in these cases.

Single best adjective. We also predicted participants' condition using only the single adjective they selected as *best* describing their evaluation of lemphurs.⁵ The discriminant analysis again revealed two functions with valence and emotionality loading most heavily on separate functions (1.00 & .99) and not on the other (.03 & -.04, respectively). Both the function associated with valence ($\chi^2(6) = 152.01, p < .001$) and emotionality ($\chi^2(2) = 28.61, p < .001$) contributed significantly. Overall, the two functions were able to predict 71.0% of participants' original condition, 46.0% better than what would be expected by chance (25.0%).⁶

⁵ Due apparently to their not understanding the instructions, six participants neglected to select a best adjective and, hence, could not be included in these analyses.

⁶ We again replicated these results with a two-way ANOVA to predict valence based on participants' best adjective. As before, more positive (negative) adjectives were used in the positive (negative) conditions and more emotional (cognitive) adjectives in the emotional (cognitive) conditions. See the Supplementary Materials for more detail.

We again assessed the ability of these single best adjectives in discriminating between positive and negative passages within the cognitive and affective passages. The implied valence of the single best adjective chosen by participants discriminated between passages for 100.0% of participants for the affective passages and 90.6% of participants for the cognitive passages.

The single best adjectives were also able to discriminate between attitude bases within a given valence. In the case of the both positive and negative passages, the discriminant analysis correctly classified 74.2% of participants.

Finally, as with using the averaging approach, when valence, extremity, and emotionality were entered into a model, the functions associated with valence and emotionality remained significant ($ps < .001$), while the function associated with extremity fell to non-significance.

DISCUSSION

By experimentally manipulating both the valence and basis of individuals' attitudes, this experiment demonstrated that our approach is able to measure the favorability of individuals' attitudes and even the attitude's source – whether it was based on cognition (an encyclopedic description of lemphurs) or affect (an emotionally evocative narrative of an individual's encounter with a lemphur) – with good accuracy. These results held both when averaging the two to five adjectives that participants selected as descriptive of their attitude as well as when considering the adjective they identified as best representing their attitude, though the averaging method had greater overall accuracy.

Building on the results from Study 1, there was additional evidence that extremity and emotionality are related but separable. Using valence, extremity, and emotionality to predict individuals' attitudes resulted in three distinct functions. As with Study 1, while the functions related to extremity and emotionality were associated, they were clearly separable. Furthermore,

we found that emotionality was a better predictor of the source of individuals' attitudes than was extremity, just as is to be expected given the focus on the manipulation of attitude basis.

This experiment also demonstrated one way in which our approach may be utilized. If researchers aim to measure individuals' attitudes toward a given object, this can be done by asking participants to select adjectives from a pre-defined list and then using the normative ratings associated with these adjectives in order to assess the attitudes and their bases.

Having validated the utility of the EL, we can now begin to ask novel and interesting questions regarding attitude basis. One such question is how emotionality and extremity separately relate to individuals' summary attitudes. Past research has indicated that emotionality does relate to summary attitudes (e.g., Abelson, Kinder, Peters, & Fiske, 1982; Lavine, Thomsen, Zanna, & Borgida, 1998; Haddock, Zanna, & Esses, 1993; Crites, Fabrigar, & Petty, 1994), but has not taken into account that emotionality and extremity are often confounded. Do attitude extremity and emotionality have separable effects when individuals seek to convey a rating that summarizes their attitude?

We can also assess cases in which individuals experience ambivalent reactions to an attitude object and ask how attitude basis informs summary judgments when positivity and negativity conflict. In other words, when people are ambivalent, does emotionality relate to how people navigate this ambivalence in either the positive or negative direction when they need to express a summary judgment? Again, past research has examined this issue, but has confounded emotionality and extremity (e.g., Lavine et al., 1998). We are in a unique position to answer these questions due to the EL's ability to dissociate between emotionality and extremity.

Study 3

In Study 3 we were not only interested in further validating our measure, but also wanted to continue to demonstrate the value of the EL for addressing important theoretical questions regarding attitudes. To do so, we turned to a more naturalistic setting in which evaluative language is common. In particular, we analyzed product reviews from Amazon.com, which contained both text reviewing the product and reviewers' overall summary rating of the product. Using this repository of data we can both validate the EL and answer theoretical questions – all within a context involving the natural use of language.

METHOD

Data

We made use of a database provided by Jindal & Liu (2008) who had used an automatized script to extract 5.9 million product reviews from Amazon.com – representing all reviews written between the years of 1996 and 2006. These reviews spanned multiple domains all the way from media (e.g., music, movies, books) to more utilitarian products (e.g., furniture, vacuum cleaners, electric toothbrushes, blenders, computers, etc) and were written by consumers interested in informing others about their experience with a given product.⁷ Each review contained a text portion where the reviewer wrote about the product, a text title for the review, and then the overall rating (out of five stars) the reviewer ultimately issued for the product.

To prepare the data for analyses, we programmed a procedure to select those reviews that contained any one of the EL adjectives and then calculate the number of times each adjective was used in a given review. We counted those adjectives that had common misspellings (e.g., “awesome”), but did not count those adjectives that were preceded by any sort of negation, either

⁷ Due to Amazon.com's product identification system, separating the different product types proved to be prohibitively difficult and therefore all product types were analyzed together.

in single words or through longer phrases (e.g., wasn't, isn't, rarely, at no time, nothing about it was, etc). Based on these criteria, 4.2 million reviews remained for subsequent analyses.⁸

RESULTS

Further validation of the emotionality dimension.

⁸ The question may be raised as to any differences between those reviews that used one of the EL adjectives and those that did not. It would appear to be difficult to not use one of the 94 adjectives when directly evaluating a product. To better understand what these reviewers were writing about, we began by examining a random sample of reviews that used the EL adjectives versus those that did not. Those reviews that did not use one of the EL adjectives tended to be more descriptive of the product and its features and more concrete in nature. These reviews tended to heavily imply an evaluation based on how the product was described, but did not actually use an EL adjective. For example, one review that was representative of this approach wrote: "These [wine savers] just do not tightly fit the neck of any wine bottle I have tried them on. They barely reach the sides and do not form a proper seal. Look elsewhere." Here the reviewer describes how the product works, or rather does not work, but does not go as far as to issue one of the EL adjectives to ascribe an evaluation. Though the reviewer did not state them, EL adjectives such as "bad," "useless," or "inferior" are heavily implied by this text. We then took a more quantitative approach in exploring the differences. First, we assessed how long, in number of words used, each review was. Reviews that used one of the EL adjectives were considerably longer ($M = 189$) than those that did not ($M = 106$; $t(5938566) = 595.94$, $p < .001$, $d = .53$), perhaps indicating greater involvement on the part of reviewers who used one of the EL adjectives. Another variable we examined was the extremity of reviewers' final product ratings. To do so, we recoded the final product rating so that any deviation from the product rating's midpoint (3 stars) was considered more extreme either in the positive direction or the negative direction. This approach resulted in a range of values from 0 to 2 representing extremity in either direction. Reviews using the EL adjectives were slightly less extreme ($M = 1.55$) compared to those that did not ($M = 1.63$; $t(5938566) = 140.54$, $p < .001$, $d = .13$). Thus, although the use of an adjective from the EL was associated with writing a lengthier review, the summary evaluation represented by the star rating was not more extreme.

To further validate the EL, we can investigate whether verbs that commonly distinguish emotion from cognition actually accompany the different adjectives within the text of the reviews. Related to past research (e.g., Mayer & Tormala, 2010), three verbs were hypothesized to be particularly good at discriminating between the different attitude bases: “feel,” “think,” and “believe.” If some of the adjectives reflect an emotional basis, the word “feel” should be particularly likely to be used in conjunction with those more emotional adjectives. The less emotional adjectives, on the other hand, should be more likely accompanied by words indicating reasoning and logic, in this case “think” or “believe.”

To test this hypothesis we extracted those reviews that contained at least one of the EL adjectives and also used just “feel” versus just “think” and/or “believe.” In other words, we did not utilize those reviews that did not use any of those three words and did not include reviews that used both “feel” and “think” and/or “believe” simultaneously. After selecting reviews based on these criteria, 1.1 million remained for analysis.

For each review, we recorded whether or not a particular adjective was used and whether the word “feel” (versus “think” and/or “believe”) was used in that same review. We therefore had a count of the number of times a particular adjective was accompanied by “feel” versus “think” and/or “believe.” If it is true that some of the adjectives in the EL reflect an emotional basis more than others do, then the normative emotionality ratings for each adjective should correlate with the proportion of times that adjective is used with “feel” versus “think” and/or “believe.” To this end, we divided the number of times an adjective was used with “feel” by the total number of times that adjective was used in our filtered sample. This calculation resulted in the proportion of the time that an adjective was used with “feel,” while one minus this proportion

represented the proportion of the time an adjective was used with “think” and/or “believe” ($M = .28$, $SD = .04$).

We then correlated this resulting proportion with the original normative emotionality ratings from Study 1. If it is the case that our adjectives correspond to a more or less emotional versus cognitive basis, then as the judged emotionality for each adjective increases the word “feel” should be used relatively more. That is exactly what we found: as the normative emotionality of our adjectives increased they were used relatively more frequently with “feel” as opposed to “think” or “believe” ($r(92) = .32$, $p < .01$). Conversely, this same result indicates that the adjectives implying a more cognitive basis were used relatively more frequently with “think” or “believe” as opposed to “feel.”

We can also ask whether this effect is specific to the implied attitude basis of the adjectives or if it also relates to their implied extremity. It is possible, for instance, that simply using more extreme adjectives may show a similar correlation with “feel” as we have seen that emotionality and extremity are related (Studies 1 & 2). The correlation of this proportion with extremity, however, was non-significant ($r(92) = .09$, $p = .39$). This finding indicates that the use of “feel” versus “think” and/or “believe” reflects the basis of an attitude rather than its extremity. This outcome makes sense as an adjective (e.g., wise) may be extremely positive, but is low on emotionality and may therefore be used less frequently with “feel.” These findings further indicate the separability of emotionality and extremity. They also serve to validate our adjectives as indicative of differential attitude bases, even within a context involving their naturalistic use.

Distinguishing univalent and ambivalent reviews.

One way we can demonstrate the value of the EL is to utilize the adjectives themselves in order to distinguish between reviews that indicate univalent attitudes and those reviews that

indicate ambivalent attitudes. In the Amazon reviews, reviewers sometimes provided an analysis that highlighted just the positive or the negative features of a product and therefore used only positive or only negative adjectives. Other times, however, reviewers highlighted both positive and negative features of the product and, hence, employed both positive and negative evaluative adjectives. The latter are likely to reflect some ambivalence about the product.

Thus, to assess ambivalence, we made a simple dichotomy between those reviews that were univalent – reviews that included only positive or only negative adjectives – and those that were ambivalent – reviews that included *both* positive and negative adjectives. On this basis, 3.4 million reviews, or 81% of the total, were classified as univalent, and 714,862 reviews, or 19% as ambivalent. This approach allowed us to ask specific questions regarding univalent and ambivalent attitudes separately. For instance, as demonstrated subsequently, we can fruitfully examine (a) how the adjectives used in univalent reviews relate to the extremity of the star ratings that the reviewers offer and (b) how the conflicting positivity and negativity in ambivalent attitudes is resolved when deciding upon and expressing a summary judgment.

To address the validity of this classification scheme, we compared the length of reviews we identified as ambivalent to the length of those identified as univalent. If it is the case that ambivalent reviewers are attempting to describe their reactions to a product that they view as having both positive and negative associations, then these reviewers may write relatively more about the product compared to those who have a more straight-forward, univalent evaluation. As hypothesized, reviews identified as ambivalent were nearly twice as long ($M = 300$ words) as those identified as univalent ($M = 166$ words, $t(4161069) = 617.70$, $p < .001$, $d = .61$).

Another way to validate this approach is to identify other words that would accompany ambivalence. In particular, we considered those words that denoted contrasting perspectives

within a review: “but,” “although,” “despite,” “in spite of,” “even though,” “nevertheless,” “on the other hand,” “though,” and “however.” We reasoned that if it is the case that the reviews we identified are ambivalent, then there should be a greater number of contrast words in the ambivalent reviews compared to the univalent reviews. Confirming this hypothesis, we found that the ambivalent reviews ($M = 2.77$) contained a greater number of contrast words than did the univalent ($M = 1.37$; $t(4161069) = 560.75$, $p < .001$, $d = .55$).

Naturally, these two correlates of reviews classified as ambivalent versus univalent – word length and greater use of contrast words – were themselves related. Longer reviews included more contrast words ($r(4161069) = .70$, $p < .001$). It is possible that simply through the course of writing longer reviews, individuals may use a greater number of contrast words. If, however, using both positive and negative adjectives within the same review is indicative of ambivalence, we should still see a greater use of contrast words in the ambivalent reviews after statistically controlling for the length of the review. Indeed, a regression equation predicting the number of contrast statements in a review simultaneously from its length ($B = 1.36$, $t(4161068) = 1878.37$, $p < .001$) and its ambivalence status ($B = .36$, $t(4161068) = 185.17$, $p < .001$) revealed that the greater use of contrast words in ambivalent reviews held true, even when controlling for the total number of words used.

A final method to validate our approach to identifying ambivalence is to use reviewers’ overall product ratings. If it is the case that these reviews are relatively more ambivalent, reviewers’ final product ratings should also be more tempered, i.e., less extreme, given that they have conflicting evaluations of the product itself. To test this hypothesis, we recoded the overall product rating to denote extremity: any deviation from the product rating’s midpoint (3 stars) was considered more extreme either in the positive direction or the negative direction. This

approach left us with a range of values from 0 to 2, with 2 representing maximum extremity for either valence. Testing the ambivalent against the univalent reviews, we found that the ambivalent reviews ($M = 1.39$) showed less extreme overall ratings compared to the univalent ($M = 1.58$; $t(4161069) = 223.30$, $p < .001$, $d = .22$).

Together, these three findings suggest that the simultaneous use of both positive and negative adjectives when providing an evaluation of an attitude object represents underlying ambivalence. For those reviews we identified as ambivalent, reviewers were more likely to contrast one viewpoint against another, write longer reviews, and express more tempered summary ratings. Based on these findings, the EL provides a useful and simple tool for identifying and studying attitudinal ambivalence.

Univalent reviews.

When they are composing their text regarding a product, reviewers typically comment on different aspects of the product itself (e.g., “the screen of the phone was good and the speaker quality was awesome”). Reviewers need to integrate these disparate aspects into a summary evaluation when they provide an overall rating for the product on the Amazon.com scale ranging from one (most negative) to five stars (most positive). The univalent reviews, on which we are currently focusing, had a mean star rating of 4.25 with a standard deviation of 1.15.

One way we can simultaneously validate the EL and provide an example of its usefulness is to examine our ability to predict reviewers’ overall summary ratings toward the different products from their naturalistic, text-based adjective use. This approach would also allow us to make novel theoretical contributions by assessing the extent to which valence, extremity, and emotionality have separable effects when individuals express a summary evaluation. Although past research has indicated that attitude basis is related to summary evaluations (e.g., Abelson,

Kinder, Peters, & Fiske, 1982; Lavine, Thomsen, Zanna, & Borgida, 1998; Haddock, Zanna, & Esses, 1993; Crites, Fabrigar, & Petty, 1994), this research has not taken into account the covariation we have noted between extremity and emotionality. Furthermore, these past studies have typically investigated a rather restricted range of attitude objects. Given the incredible variety of products the Amazon reviews provide us, we can investigate the effects of extremity and attitude basis with greater confidence that any observed findings are not limited to a small number of attitude objects.

To begin our analyses, we quantified the text of each univalent review by creating a weighted average relevant to each of the variables of interest. Specifically, each review was analyzed for how many times each of the EL adjectives was used, including if an adjective was used more than once in a given review. These counts were then multiplied by each adjective's normative ratings (valence, extremity, and emotionality) as indexed in the EL. Finally, this resulting number was divided by the number of times one of the EL adjectives appeared in that review. Take the following sentences as an example: "This product was amazing. I thought the camera lens was nice and the design was appealing. The photos were simply amazing." The underlined adjectives have normative emotionality ratings of 6.59, 5.53, and 5.38, respectively. The emotionality index for this review would therefore be: $(2 * 6.59 + 1 * 5.53 + 1 * 5.38) / 4 = 6.02$ (see Dodds & Danforth, 2010, for a similar calculation). This same equation was also used with both the valence and extremity implied by reviewers' adjective use. This approach allowed us to use all possible adjectives in the text and to account for those adjectives that were used by the reviewer multiple times, thereby suggesting that such an adjective may be all the more indicative of a reviewer's attitude toward the product.

As an initial step, we assessed the associations between extremity and emotionality within these univalent reviews. This correlation is all the more interesting as we are now assessing the actual use of the adjectives in natural text. It is entirely possible that not all adjectives are used as frequently as one another in a naturalistic context. This differential frequency could result in a correlation that differs from that obtained in the normative ratings examined in Study 1. The approach in Study 1 gave equal “weight” to each adjective. In the present context, each adjective’s impact upon the correlation varies as a function of the frequency with which it appears in the reviews. What we found, however, was a replication of the effect in Study 1 as emotionality and extremity showed a moderate association across reviews ($r(3446208) = .64, p < .001$). Again, this result demonstrates the relation and separability of emotionality and extremity, even in natural text.

Turning our attention now to our primary analysis predicting reviewers’ summary ratings, while it seems evident that reviewers would be unlikely to use positive adjectives when reviewing a product toward which they have developed a negative attitude – and vice-versa – it remains to be seen whether the implied extremity of the adjectives that reviewers use also contributes to the prediction of their summary ratings. It will be even more telling if the implied attitude basis (i.e., the emotionality of the adjectives that are used) also plays a role above-and-beyond valence and extremity. The effects of extremity and emotionality should vary as a function of valence: more implied extremity or emotionality in a positive adjective should predict more extreme positive ratings while these same variables should predict more extreme negative ratings for negative adjectives.

Thus, we used valence, extremity, emotionality, and two two-way interactions, valence by extremity and valence by emotionality, to predict the summary ratings. Although both the

extremity and emotionality variables could be entered into the model without alteration, the valence variable required greater consideration. Specifically, because extremity was directly calculated from the valence ratings provided by the judges in Study 1, these variables begin to overlap as soon as the direction of the extremity is identified as either positive or negative. Due to this redundancy, the valence variable was dummy coded: if the valence score for any given review was greater than the midpoint of the original normative scale (i.e., greater than 4.5), then the review was categorized as positive and coded as 1; otherwise, it was categorized as negative and coded as 0. In essence, then, the valence variable acts to define whether the implied extremity and emotionality is in terms of positivity or negativity.

As predicted, the two two-way interactions were significant in a regression analysis predicting star ratings. The valence by extremity ($B = .29$, $t(3446209) = 119.72$, $p < .001$) and valence by emotionality ($B = .13$, $t(3446209) = 58.10$, $p < .001$) interactions indicated that as the implied extremity or emotionality of the adjectives within reviewers' text increased, the extremity of their summary evaluation also increased (see Figure 1).⁹ All simple slopes were significant ($ps < .001$).¹⁰ In essence, increased extremity and emotionality in the text of the

⁹ The analysis also revealed a significant three-way interaction between valence, extremity, and emotionality ($B = -.02$, $t(3446207) = 21.31$, $p < .001$). However, as is apparent from the very small B , the interaction accounted for negligible variance. Moreover, graphing this interaction failed to reveal any meaningful difference from what was evident on the basis of just the two two-way interactions.

¹⁰ As the reader may have noticed from the mean of the product ratings, Amazon.com reviews are heavily skewed; there are many more positive than negative reviews. Due to this skew and possible concerns of non-normality, we wanted to make sure our results were robust across methods of analysis. We proceeded by taking just those reviews that were maximally negative (1-star; $n = 204,651$) and an equivalently-sized random sample of those that were maximally positive (5-stars; $n = 204,739$). We then repeated the original analyses using logistic regression to predict

reviews predicted greater polarization of summary judgments. The effect of extremity was such that the difference between the reviews that used the least extreme positive adjectives to those with the most extreme positive adjectives was over one full star rating. Increasingly extreme negative reviews yielded a change of about 1/4 of a star rating. A similar pattern was found with emotionality such that a shift from the reviews that used the least to the most emotional adjectives, despite holding extremity constant, represented a change of about 1/3 of a star rating.¹¹ Past research has shown that consumers' purchasing decisions are influenced by Amazon's product rating system. Increases or decreases in average star ratings for a particular product on Amazon.com are directly related to later increases or decreases in sales, even while holding changes in sales at other related websites constant (Chevalier & Mayzlin, 2006; see also Luca, 2011). Thus, the differences in star ratings we note here can affect which product consumers purchase and, hence, the revenue of a business.

whether a review was positive or negative. All analyses were fully replicated using this alternative method ($p < .001$). It therefore appears that the skewness of the data did not adversely affect the results using all reviews.

¹¹ We also conceptually replicated these results using the titles of the reviews. Titles of reviews appear to largely be summaries of the body of the text. Thus, instead of relying solely on star ratings, we can predict the average valence of the title of the reviews from the body of the reviews. To begin, we identified those univalent reviews that also used any of the EL adjectives in the title and were left with 899,199 reviews. We then predicted the average valence implied by the adjectives used in the title. Replicating the results predicting the star ratings, both the valence by extremity ($B = .63, t(899193) = 81.94, p < .001$) and valence by emotionality ($B = .24, t(899193) = 32.65, p < .001$) interactions were significant. As with the star ratings, these interactions indicated that as the implied extremity or emotionality within reviewers' body text increased, the extremity of their title evaluation also increased.

In a manner that conceptually parallels the approach pursued in Study 2 with participants' designation of a single-best adjective as a summary of their evaluation, we also used adjectives in the title of reviews to predict star ratings. As detailed in the Supplementary Materials, this analysis also replicated the findings reported above.

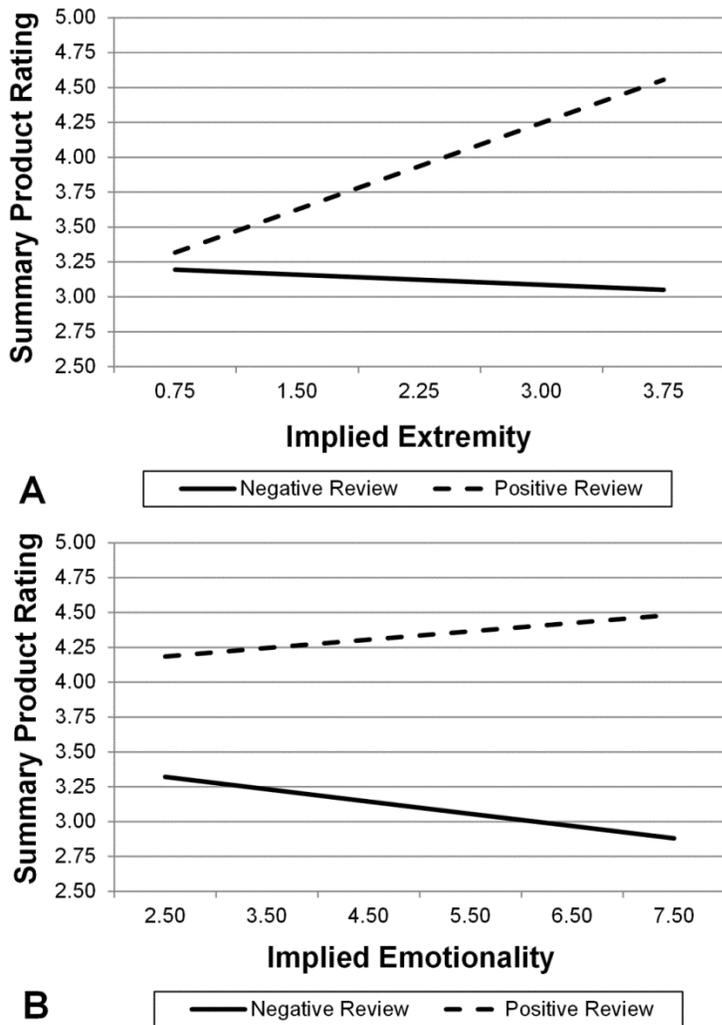


Figure 1 – A) Regression lines relating the implied extremity of the adjectives in a review as a function of valence. B) Regression lines relating the implied emotionality of the adjectives in a review as a function of valence. Values on the x-axis range from roughly the minimum to the maximum values within the sample.

As can be seen in Figure 1, although it appears implied emotionality has a similar effect on each valence, extremity appears to be particularly important for positive reviews. One point on the regression line to take notice of is the positive review at low extremity (see Figure 1, Panel A). This point estimate appears to be quite low compared to the results for positive emotionality, for example, while all the other point estimates are fairly similar. One reason for this might be the way products are released into the market. It is very rare for companies to release products that are completely terrible given that each product must usually pass a number of requirements to even reach the marketplace. This may explain why there are fewer outright negative reviews on Amazon.com. It would also help to explain the lower point estimate for less extreme positive reviews. In essence, reviewers are using words like “acceptable” or “adequate” when they are reviewing these products, both of which are low on positive extremity. For products that have made it beyond companies’ initial testing, these reviews may actually indicate quite lukewarm evaluations at best and therefore may reflect particularly low summary ratings.

Overall, these findings demonstrate the ability of the normative valence, extremity, and emotionality ratings from Study 1 to predict meaningful outcomes in an entirely different context, thereby further validating the EL. These findings also establish an important theoretical point: attitude basis matters above-and-beyond attitude extremity. This is the first demonstration that emotionality predicts summary evaluations even while simultaneously holding extremity constant. While past studies measuring attitude basis have shown that attitude emotionality predicts summary evaluations, these studies have not had the ability to assess the distinct influences of both emotionality and extremity. As we have demonstrated, it is important to assess extremity as it is associated with emotionality and also plays an important role in predicting summary evaluations. Our approach gives us the unique flexibility to control for the effect of

extremity, thereby allowing us to better isolate the effect of attitude basis apart from attitude extremity. These results are also consistent with the idea that attitude emotionality can influence judgments in ways that are not directly related to extremity (Fazio, 1995).

Ambivalent reviews.

As mentioned above, we also identified ambivalent reviews on the basis of the reviewers using both positive and negative adjectives. This approach resulted in 714,862 reviews, 19% of the total; they received a mean summary rating of 3.70 with a standard deviation of 1.40.

Given that these reviews reflect ambivalence on the part of reviewers, we can also use these data to further demonstrate the EL's usefulness and its ability to contribute to social psychological theory regarding attitudes. When composing their reviews, individuals who use both positive and negative adjectives are essentially indicating that some aspects of the product are positive, but that others are negative. They therefore are required to come to some sort of resolution if they wish to offer a single summary evaluation. To date, there has been relatively little work assessing how different aspects of attitudes (e.g., their basis) are related to how individuals resolve their ambivalent reactions when expressing a single summary evaluation. For instance, when positive and negative valence conflict, how does having one valence based more on affect than the other relate to the individual's decision to communicate a particular summary evaluation? If such a role for affectively-based valence exists, does it hold above-and-beyond the extremity of that valence? Finally, what role does focusing on one valence more than another play in these relationships? We are able to address these questions with the EL.

To our knowledge there is only one finding assessing this issue in the literature (Lavine et al., 1998). In this particular study, the researchers analyzed a pre-existing database of attitudes toward presidential candidates for the 1980 to 1992 presidential elections. Respondents had been

asked to indicate whether a candidate had ever evoked within them a particular emotion or not (e.g., hopeful, proud, angry, afraid), to what extent a candidate exhibited a given trait (e.g., moral, knowledgeable, dishonest, weak) – “cognition” – and their overall summary evaluation of each candidate. The researchers found that when affect and cognition conflicted, affect tended to better predict overall evaluations). As noted earlier, these results are ambiguous as emotionality and extremity tend to be confounded. Furthermore, this past research limited its consideration of ambivalence to conflicting affective and cognitive reactions. In contrast, the EL provides us with the added benefit of allowing ambivalence to be defined simply as the expression of both positivity and negativity, which could assume any form (e.g., negative affect versus positive affect, negative cognition versus positive cognition, etc). In addition, this research focused on a single, special type of attitude object (people), who we know to be judged in ways that are quite different than other attitude objects (e.g., Skowronski & Carlston, 1989). The extent to which other attitude objects are assessed similarly is an interesting and open question. Finally, because reviewers are free to express their evaluation in any way they choose, we also have the unique ability to assess which valence reviewers are focusing on by their own accord.

As with univalent reviews, we sought to predict reviewers’ overall summary evaluation of the products from the valence, extremity, and emotionality implied within their reviews. For the ambivalent reviews, however, we required additional variables as the reviews communicate both positive and negative extremity and emotionality. Indeed, our primary question was how ambivalence is resolved into a single summary judgment when positivity and negativity conflict.

To create the needed variables, we began by categorizing adjectives as either positive or negative as indicated above. For both extremity and emotionality we then created a weighted average – using the same approach as we did for univalent reviews – for all positive adjectives

within the review and then a separate weighted average for all the negative adjectives. This resulted in four variables for each review: positive and negative extremity as well as positive and negative emotionality. We were additionally interested in how focusing on one valence versus the other mattered. For example, did it matter if a reviewer used three positive adjectives and one negative adjective – thereby indicating greater focus on the positive valence – compared to a reviewer who used just two positive adjectives and one negative? Furthermore, did this focus variable interact with our other variables of interest? It is possible, for instance, that giving greater weight to positivity but having negativity based more in emotionality may lead individuals to navigate their ambivalence differently than if they give equal focus to each valence. To that end we created two additional variables counting the number of positive and negative adjectives used in each review as a measure of valence focus.

As our question concerned how people go about constructing summary evaluations when positivity and negativity conflict, we created three difference variables between positivity and negativity from the six variables we generated previously. Specifically, we subtracted negative extremity from positive extremity, negative emotionality from positive emotionality, and the number of negative adjectives from the number of positive adjectives.

After creating these three difference variables, we standardized and then entered them into a regression equation, with their interactions, to predict reviewers' summary product ratings. The effects of the valence focus ($B = .27$, $t(714854) = 153.34$, $p < .001$), extremity ($B = .22$, $t(714854) = 106.76$, $p < .001$), and emotionality ($B = .05$, $t(714854) = 25.57$, $p < .001$) difference variables were all significant. Ratings were more positive when the reviewer used more positive than negative adjectives. They also were more positive when the positive adjectives that were used implied greater extremity than the extremity implied by the negative adjectives that were

used. Thus, ambivalent reactions were resolved in the direction of the valence that received more frequent mention and the valence associated with the greater extremity.

Even more theoretically interesting, however, was that these results were also true of emotionality: the more emotionally-based either valence was, the more likely this valence was to dominate when reviewers provided a final summary evaluation. This effect of emotionality was seen over-and-above the focus on each valence and its implied extremity. These results become particularly noteworthy when one considers the words that are being used. For example, although “useful” and “enjoyable” are fairly similar in their implied positive extremity (~ 2.82), they differ considerably in their implied emotional basis (3.14 and 6.77, respectively). Thus, an object that an individual describes using the more cognitive adjective “useful” may essentially lose the competition when the individual also describes it with a more emotional, negative adjective of the same extremity (e.g., “ridiculous,” which has an implied emotionality rating of 5.11). However, if “enjoyable” were to be pitted against “ridiculous,” the greater emotionality of “enjoyable” would be enough to shift the summary evaluation in a more positive direction.

An extremity by emotionality interaction also was evident ($B = -.06$, $t(714854) = 39.82$, $p < .001$; see Figure 2). Although the valence with greater implied extremity dominates, this effect is attenuated when the positive valence is associated with greater emotionality. It appears that more positive emotionality benefits an object associated with greater negative extremity (i.e., 1 SD below the mean for differential extremity; $B = .11$, $t(714858) = 44.44$, $p < .001$). However, more negative emotionality did not similarly degrade an object associated with positive extremity (i.e., 1 SD above the mean; $B = -.004$, $t(714858) = 1.63$, $p = .10$).¹²

¹² As with the univalent reviews, we again had the opportunity to conceptually replicate our results using the titles of the ambivalent reviews. When predicting the valence of the title of the reviews, all the effects reported for the body



Figure 2 – Regression lines relating the implied positive versus negative emotionality of the adjectives in a review as a function of extremity. Plotted points are one standard deviation above and below the mean.

of ambivalent reviews were replicated. Furthermore, we replicated the two-way interaction between differential extremity and emotionality. As before, more positive emotionality benefitted an object associated with greater negative extremity (extremity 1 SD below the mean; $B = .27$, $t(170458) = 31.40$, $p < .001$). This time, however, greater negative emotionality significantly degraded an object associated with positive extremity, albeit to a lesser degree than the other simple slope (extremity 1 SD above the mean; $B = .07$, $t(170458) = 7.76$, $p < .001$). See the Supplementary Materials for more detail.

DISCUSSION

In Study 3 we looked to further validate our approach as well as demonstrate its ability to contribute to practical and theoretical issues, this time outside of the lab using natural text. To begin, using those Amazon.com reviews that included at least one of the adjectives from the EL, we showed that adjectives that implied a greater emotional basis were used relatively more with the word “feel” while adjectives that implied a less emotional basis were used relatively more with the words “think” or “believe.” This further validates the EL adjectives by demonstrating that those adjectives that implied an emotional basis are indeed accompanied by words indicating the reviewer’s attitude was more based on affect while those that implied a less emotional basis were accompanied by words indicating the reviewer’s attitude was more based on cognition.

We then demonstrated the EL’s usefulness in identifying and then analyzing univalent and ambivalent reviews. We identified reviews as ambivalent if they used both the EL’s positive and negative adjectives. To validate this classification, we then compared these reviews to those we had identified as univalent based on their using just positive or just negative adjectives. Ambivalent reviews were more likely to contain words that denoted contrasting opinions than did univalent reviews, were longer than univalent reviews, and were associated with less extreme, more tempered summary ratings. All of these characteristics further validate our approach in identifying ambivalence within the current sample, but also point to its usefulness for future studies investigating ambivalence.

As a next step we then analyzed just those reviews that expressed a univalent attitude (i.e., those reviews that used just positive or just negative adjectives). We found that the valence, extremity, and emotionality ratings implied by the reviewers’ adjective use in their main text predicted their summary evaluation of each product in meaningful ways. Specifically, the more

reviewers used adjectives in the text that implied extremity or emotionality, the more polarized their summary ratings were of the products they were reviewing. We obtained results using just the titles that reviewers provided for their reviews, thereby increasing confidence in the original results and extending them to shorter-length text (see Footnote 11 and the Supplementary Materials). Taken together, these results further validate our general approach as implied valence, extremity, and emotionality all predicted reviewers' summary evaluations. They also provide an initial demonstration of the EL's practical value as it provided sensible results using natural text outside of the lab. This means that attitudes can be fruitfully measured using available natural-text repositories where new and interesting hypotheses can potentially be tested. Finally, the present findings support the theoretical proposition that emotionality and extremity are separable and also that attitude basis matters above-and-beyond the effects of extremity.

As a final step, we asked an additional theoretical question of our data: when their reactions include both positive and negative aspects, how do individuals resolve this conflict into a final summary evaluation? Our results indicate that, on average, valence focus, extremity, and emotionality all relate to the overall summary that reviewers communicate. Whichever valence is mentioned more frequently, and whichever implies the greater extremity or emotionality is the valence that is more likely to dominate when individuals need to resolve their ambivalence in order to express a single summary evaluation. The effect of emotionality is particularly noteworthy given that the model controlled for both valence focus and extremity. We demonstrated this pattern of findings both using the body of the text to predict final star ratings as well as with analyses of the titles of the reviews (see Footnote 12 and the Supplementary Materials). These results again demonstrate the separability of extremity and emotionality.

Furthermore, as opposed to past research on ambivalence resolution and attitude basis which focused solely on judgments regarding people, we showed that the effects of extremity and emotionality occur across an extremely wide variety of attitude objects.

An additional point we wish to make is that although we separated the reviews into those that were univalent and those that were ambivalent in order to address distinct theoretical questions, researchers wishing to use the EL to analyze natural text have a multitude of possible ways to approach their data. It certainly is not necessary to treat the two classes of reviews separately. Values can be imputed for all the adjectives that appear in an ambivalent review and then averaged to form a single implied valence and implied emotionality score as we did for the univalent reviews. Doing so with the current data produced outcomes like those we had obtained with the univalent reviews. The greater both the average extremity and emotionality of the adjectives in the reviews, the more polarized their summary star ratings tended to be.¹³ Thus researchers have the ability to consider univalent and ambivalent reviews simultaneously if that is more suitable to their aims.

GENERAL DISCUSSION

The aims of the present research were to examine the implications of the multitude of adjectives that language provides for expressing evaluations and to then leverage this language to make novel theoretical contributions. The research findings indicate that the adjectives individuals use when describing their reactions to a given object provide important information regarding their attitudes toward that object. As such, this research validates an approach to

¹³ As with just the univalent reviews, when using all reviews we found two-way interactions between valence and extremity ($B = .26, t(4161065) = 120.89, p < .001$) and between valence and emotionality ($B = .11, t(4161065) = 54.00, p < .001$).

measuring different aspects of attitudes through language use. In particular, we provided a tool – the Evaluative Lexicon – that allows researchers to index the valence, extremity, and basis of individuals’ attitudes through their use of adjectives.

In Study 1, we created a broad list of adjectives that fit a number of important prerequisites. After creating this list we asked judges to rate the adjectives on their implied valence and implied emotionality. These normative ratings provide the substance of the EL for use in future research.

In Study 2, we validated our approach by experimentally creating attitudes within a laboratory setting and then having participants select those adjectives that best described their evaluation. We were able to predict the basis of the newly-created attitudes at a very high rate from individuals’ adjective use. This outcome was true both for valence as well as the actual source, or basis, of the attitude – whether it was an encyclopedic-like description of the attributes of the attitude object or an emotional narrative passage involving the target object. This experiment builds on Study 1 by demonstrating that not only can individuals rate the adjectives meaningfully, but that they can also use them to express their attitudes and their attitudes’ bases.

In Study 3, we sought to validate our approach in a different setting and then to further demonstrate its ability to advance social psychological theory regarding attitudes using a natural-text repository of product reviews from Amazon.com. Even with individuals outside of the lab who were unconstrained to selecting a limited number of adjectives, our results demonstrated that the normative ratings obtained from judges in Study 1 were meaningfully associated with a number of important phenomena. For instance, we found that the more emotional adjectives were associated with reviewers using the word “feel” while the less emotional, cognitive adjectives were used with “think” and “believe.”

Study 3 also allowed us to address novel theoretical questions about attitudes. We showed that an attitude's basis matters for both univalent and ambivalent attitudes above-and-beyond attitude extremity. Univalent attitudes show more polarized summary evaluations when based more on emotion, over and above the effects of extremity. When reviewers expressed ambivalent reactions to a product, the valence associated with greater emotionality tended to govern the summary evaluation that the reviewer communicated. Thus, when positivity and negativity conflicted, the valence based more on emotionality won the day. Again, this effect of emotionality held over-and-above other important constructs such as valence focus and extremity, and it was present across an extremely wide variety of attitude objects. Past research has not accounted for the simultaneous effects of emotionality and extremity, has limited ambivalence to occur between affect and cognition, and has tended to focus on evaluations of people, who we know to be a special kind of attitude object (e.g., Skowronski & Carlston, 1989).

Throughout all three studies, we have thus found empirical support for the hypothesis that attitude extremity and emotionality, although related, are separable. For instance, Study 2 revealed that the source of the attitudes we created in the lab was better predicted by emotionality rather than extremity. Study 3, on the other hand, demonstrated that both extremity and emotionality can have simultaneous effects. This article begins the process of distinguishing between extremity and emotionality. It will be the task of future research, then, to further illuminate the effects of attitude basis apart from extremity and vice-versa. It is our hope that our initial demonstration here will spur further interest in this vein.

Though other tools exist to quantify text, our approach is considerably different than those that currently exist. Of these, the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) system is the most well-known and popular in psychology. One major

difference lies with the original purpose of LIWC. Specifically, LIWC was created to analyze the positive and negative emotion words that individuals wrote (e.g., “distraught,” “upset,” “happy”) and how these related to future well-being (e.g., Pennebaker, 1993). As such, LIWC tends to concentrate on expressed emotions rather than individuals’ description and evaluations of different objects. Given these different aims, LIWC does not contain about one-third (31) of our adjectives. A *t*-test of those words that LIWC does ($M_{\text{extremity}} = 3.19$; $M_{\text{emotionality}} = 6.57$) and does not include ($M_{\text{extremity}} = 2.74$; $M_{\text{emotionality}} = 5.71$) revealed that those words that are not included in LIWC tend to imply less extremity and emotionality (e.g., “beneficial,” “appealing,” “mediocre,” “harmful;” $ps < .01$). These results make sense given LIWC’s aim and history of concentrating on expressed emotionality and not on attitudes and their different aspects. A second major difference between our approach and LIWC is that LIWC quantifies text using a count of how frequently particular words are used. This frequency calculation is then associated with outcomes of interest. Our approach, on the other hand, began with existing theory on attitude bases and utilizes normative ratings concerning the connotations of each evaluative adjective. As such, these two approaches are very different.

The current research also extends existing methods of measuring attitudes and their bases and we believe our approach is characterized by a number of advantages. First, the EL does not require individuals to introspect separately on either the characteristics of the object or their feelings toward it. Past methods required that participants explicitly focus on the different possible bases (i.e., cognitive versus affective) one-by-one. Instead, we are able to ask individuals a single question, “which of the following adjectives best describe your attitude toward this object?”, and the respondents have the freedom to simply use those adjectives that

best represent their evaluations, without the need to consider a directive to focus on either their affective or cognitive reactions.

Second, our approach is less susceptible to the possibility of missing data. When using the open-ended approach to measuring attitude bases, for instance, researchers can lose data when participants are unable to list any reactions to the given attitude object.

Third, our approach enhances the ease with which individuals can complete the measure and therefore increases efficiency in terms of time for researchers. Indeed, participants need not list their own reactions nor complete multiple scales for the same attitude object. Instead, they can select even just a single adjective and this adjective can, in turn, represent multiple variables.

Fourth, our approach allows researchers to consider a *given* attitude and assess its basis. Both the more open-ended and closed approaches to measuring attitude bases have required separate subscales to measure the different bases and have relied on “relativizing” the attitude and its basis to other attitudes that are also being measured (e.g., Crites et al., 1994; Eagly et al. 1994; Haddock et al., 1993). This relativizing can proceed in a number of different ways but often is done by either regressing the subscales onto individuals’ summary evaluations or z-scoring the subscales and computing absolute difference scores (see Eagly et al., 1994 and See, Petty, & Fabrigar, 2008 for examples). Thus, researchers can infer only that an individual tended to have an attitude that was based relatively more on affect (or cognition) than another participant or, if a within-subject orientation was pursued, that an attitude toward a given object was based relatively more on affect (or cognition) compared to the individual’s other attitudes (e.g., Huskinson & Haddock, 2004; See et al., 2008). The approaches make it difficult for a researcher to measure a single attitude from a participant and know whether that attitude is based more on affect or cognition. We avoid these issues by imputing our knowledge of the

connotations of the adjectives that individuals use to describe their attitudes. We do not need to pursue any “relativizing” calculations, but instead can infer the attitude basis from the position of the adjective within the EL. When a person uses an adjective such as “awesome” to describe their evaluation of an object, the EL informs us that the evaluation implies emotionality.

Finally, as demonstrated in Study 3, our method allows for researchers to test practical and theoretical questions both within a laboratory setting and through natural text use within a diversity of settings.

The present research also illustrates the power of language in regard to attitudes. We find it remarkable that the English language’s adjectives allow individuals to communicate such nuances regarding their evaluations. Not only do people recognize these nuances when explicitly asked to analyze the adjectives (Study 1), but they can use them fruitfully in the lab (Study 2) as well as in natural text (Study 3). Although our focus was on putting forth the EL and demonstrating its ability to measure the separate influences of valence, extremity, and emotionality, it is easy to envision other ways these adjectives could be quantified in the future, e.g., the extent to which an adjective seems related to issues of morality (e.g., Skitka, Bauman, & Sargis, 2005), the extent to which it implies active approach-avoidance versus passive acceptance, or the extent to which it suggests a past, current, or future orientation. It is our hope that bringing researchers’ attention to this opportunity may generate further progress in using language to measure different aspects of attitudes and thus expand the lexicon.

Future Directions

Given that one of this paper’s main aims was to develop a tool to measure attitudes across a number of different domains, we believe that the results provided here represent a starting point for a wealth of future research. One interesting branch of future research would be to see how

others perceive individuals or text that uses these adjectives. For example, if individuals use more emotional adjectives to describe different attitude objects, is this apparent to people interacting with or reading statements by these individuals? What effects does this have on the recipient of the information? In terms of products like those in our Amazon.com study, it is possible that less emotional and more extreme adjective use may increase the perceived usefulness of the information. These kinds of adjectives may represent more thoughtful, attribute-based evaluations of products (e.g., use of the words “wise” or “beneficial”) compared to their emotional counterparts, which may then be perceived as more subjective and idiosyncratic.

In terms of attitudinal ambivalence, while we showed in Study 3 that the number of positive and negative adjectives used, their implied extremity, and their implied emotionality all matter when resolving ambivalence, it is possible this is mostly true under low time constraint. In other words, when writing their product reviews, Amazon reviewers have ample time to come to an understanding of how to resolve their ambivalence. It may be possible, however, that when they are required to make a judgment quickly, individuals may rely on certain aspects of their attitudes to a greater extent than others (e.g., relying more on the emotional basis and less on extremity).

These and many other questions can be fruitfully addressed in future research employing the Evaluative Lexicon. What we find particularly exciting is the ability of the Evaluative Lexicon to test new hypotheses in both laboratory research and natural-text repositories across a wide range of domains.

REFERENCES

- Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology*, 42(4), 619–630.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.
- Crites, S. L., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin*, 20(6), 619–634.
- Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4), 441–456.
- Eagly, A. H., Mladinic, A., & Otto, S. (1994). Cognitive and affective bases of attitudes toward social groups and social policies. *Journal of Experimental Social Psychology*, 30(2), 113–137.
- Fabrigar, L. R., & Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Personality and Social Psychology Bulletin*, 25(3), 363–381.
- Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude Strength: Antecedents and Consequences* (pp. 247-282). Mahwah, NJ, Erlbaum.
- Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, 65(6), 1105–1118.

- Haddock, G., & Zanna, M. P. (1998). On the use of open-ended measures to assess attitudinal components. *British Journal of Social Psychology*, *37*(2), 129–149.
- Holtgraves, T. (2010). Social Psychology and Language: Words, Utterances, and Conversations. In *Handbook of Social Psychology* (pp. 1386–1422). New Jersey: John Wiley & Sons, Inc.
- Huskinson, T. L. ., & Haddock, G. (2004). Individual differences in attitude structure: Variance in the chronic reliance on affective and cognitive information. *Journal of Experimental Social Psychology*, *40*(1), 82–90.
- Jindal, N., & Liu, B. (2008, February). Opinion spam and analysis. In *Proceedings of the international conference on Web Search and Web Data Mining* (pp. 219-230).
- Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp.com (Working Paper 12-016). Harvard Business School, Boston, MA. Retrieved April 25, 2014, from <http://hbswk.hbs.edu/item/6833.html>.
- Mayer, N. D., & Tormala, Z. L. (2010). “Think” versus “feel” framing effects in persuasion. *Personality and Social Psychology Bulletin*, *36*(4), 443–454.
- Norman, R. (1975). Affective-cognitive consistency, attitudes, conformity, and behavior. *Journal of Personality and Social Psychology*, *32*(1), 83–91.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). LIWC2007: Linguistic inquiry and word count. *Austin, Texas: LIWC.net*.
- Pennebaker, J. W. (1993). Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour Research and Therapy*, *31*(6), 539–548.
- See, Y. H. M., Petty, R. E., & Fabrigar, L. R. (2008). Affective and cognitive meta-bases of attitudes. *Journal of Personality and Social Psychology*, *94*(6), 938–955.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895–917.

Van den Berg, H., Manstead, A. S. R., van der Pligt, J., & Wigboldus, D. H. J. (2006). The impact of affective and cognitive focus on attitude formation. *Journal of Experimental Social Psychology*, 42(3), 373–379.

Appendix

Table A1. Means and standard deviations of the averaged implied valence, extremity, and emotionality for each of the 94 adjectives. Arranged from most positive to most negative.

	Valence		Extremity		Emotionality	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Magnificent*	8.71	0.60	4.21	0.60	6.80	2.18
Excellent	8.63	0.64	4.13	0.64	5.17	2.21
Awesome*	8.59	0.65	4.09	0.65	6.78	2.07
Fantastic*	8.57	0.70	4.07	0.70	6.64	2.04
Perfect	8.53	1.28	4.16	0.73	4.72	2.61
Amazing*	8.53	0.72	4.03	0.72	6.59	2.32
Outstanding	8.44	1.21	4.07	0.63	5.92	2.33
Wonderful	8.41	0.78	3.91	0.78	6.98	1.83
Fabulous	8.16	0.99	3.66	0.99	6.59	1.90
Lovable*	8.15	0.89	3.65	0.89	7.19	1.88
Great	8.09	0.89	3.59	0.89	5.52	2.29
Very good	7.96	0.94	3.47	0.88	4.83	2.33
Wise*	7.93	0.98	3.43	0.98	3.84	2.64
Terrific	7.88	1.83	3.72	0.96	6.06	2.35
Joyful	7.84	1.02	3.34	1.02	7.61	1.71
Exciting*	7.78	1.18	3.29	1.14	7.33	1.59
Smart	7.76	1.17	3.34	0.94	2.89	2.25
Positive	7.76	1.19	3.26	1.19	4.98	2.36
Delightful*	7.75	1.26	3.32	1.05	7.27	1.43
Valuable*	7.68	1.00	3.18	1.00	3.98	2.46
Attractive	7.66	1.02	3.16	1.02	5.52	2.50
Healthy	7.53	1.28	3.10	1.08	2.92	2.39
Cheerful	7.53	1.18	3.03	1.18	6.98	1.78
Beneficial*	7.51	1.18	3.03	1.14	3.55	2.46
Enjoyable*	7.47	0.98	2.97	0.98	6.77	1.81
Desirable*	7.43	1.08	2.93	1.08	6.59	1.92
Pro	7.41	1.44	3.00	1.24	2.78	2.31
Helpful	7.37	1.14	2.87	1.14	4.17	2.35
Favorable	7.32	1.09	2.82	1.09	4.84	2.50
Superior	7.29	1.84	3.10	1.24	4.36	2.43
Pleasant*	7.26	0.97	2.76	0.97	5.88	2.16

Relaxing*	7.25	1.18	2.75	1.18	5.16	2.95
Worthwhile*	7.16	1.24	2.68	1.21	4.50	2.37
Likable	7.16	1.05	2.66	1.05	6.05	2.18
Appealing*	7.15	1.41	2.78	1.12	5.38	2.45
Useful*	7.13	1.18	2.66	1.11	3.14	2.56
Good	7.09	1.19	2.60	1.16	4.69	2.03
Wholesome	7.00	1.41	2.63	1.14	3.78	2.56
Calming*	6.91	1.27	2.43	1.24	5.70	2.56
Commendable*	6.87	1.60	2.60	1.17	4.27	2.58
Nice	6.87	1.09	2.38	1.06	5.53	2.23
Safe	6.84	1.78	2.59	1.39	4.00	2.44
Agreeable	6.53	1.23	2.07	1.15	4.39	2.51
Reasonable*	6.15	1.45	1.79	1.26	3.41	2.34
Acceptable*	6.06	1.38	1.69	1.21	3.38	2.47
Satisfactory	5.94	1.46	1.60	1.28	3.52	2.42
Okay	5.34	1.07	1.07	0.83	3.48	2.45
Adequate	5.13	1.35	1.10	0.99	2.83	2.13
Neutral	4.65	0.93	0.63	0.69	2.50	2.25
Average	4.32	0.94	0.65	0.70	2.45	2.14
Tolerable	4.31	1.70	1.25	1.16	3.95	2.24
Mediocre	3.29	1.50	1.46	1.25	3.78	2.21
Questionable*	2.74	1.09	1.81	1.01	3.56	2.22
Imperfect	2.54	1.44	2.15	1.13	3.14	2.38
Objectionable*	2.41	1.35	2.15	1.26	5.05	2.50
Boring	2.25	1.36	2.29	1.29	4.78	2.96
Foolish*	1.97	1.12	2.54	1.08	4.59	2.59
Ridiculous	1.91	1.17	2.60	1.13	5.11	2.32
Sorrowful*	1.85	1.43	2.71	1.31	7.19	2.30
Inappropriate	1.71	1.07	2.81	1.03	4.28	2.64
Troublesome*	1.68	1.09	2.82	1.09	4.70	2.41
Dislikable	1.65	1.13	2.85	1.13	5.30	2.49
Unhealthy*	1.54	1.14	2.96	1.14	3.00	2.49
Upsetting*	1.54	1.15	2.96	1.15	6.80	2.42
Saddening*	1.53	1.09	2.97	1.09	6.78	2.45
Inferior*	1.50	1.47	3.13	1.16	4.17	2.55
Con*	1.43	1.34	3.16	1.11	3.28	2.63
Unsafe*	1.43	1.18	3.07	1.18	3.92	2.55
Annoying*	1.35	1.02	3.15	1.02	6.42	2.35
Offensive	1.29	0.98	3.21	0.98	6.63	2.19
Angering	1.28	1.06	3.22	1.06	6.81	2.54
Irritating*	1.26	1.09	3.24	1.09	6.70	1.94
Stupid	1.24	1.22	3.26	1.22	4.45	2.74
Bad	1.22	1.21	3.29	1.17	4.88	2.50

Frightening	1.22	1.06	3.28	1.06	7.34	1.87
Harmful*	1.21	1.54	3.50	0.98	4.45	2.73
Dangerous*	1.18	1.18	3.34	1.14	5.05	2.52
Negative*	1.12	1.15	3.38	1.15	4.94	2.42
Undesirable	1.12	0.95	3.38	0.95	5.27	2.48
Disturbing	0.96	0.90	3.54	0.90	6.59	2.40
Appalling	0.91	1.10	3.60	1.05	6.84	2.28
Depressing	0.91	0.86	3.59	0.86	7.25	2.17
Useless*	0.84	0.99	3.66	0.99	4.25	2.65
Disgusting	0.74	0.91	3.76	0.91	6.78	2.16
Terrifying*	0.72	0.93	3.78	0.93	7.50	1.97
Sickening	0.71	1.09	3.84	0.92	6.52	2.23
Horrible	0.68	1.25	3.96	0.72	7.06	1.96
Awful*	0.66	0.87	3.84	0.87	6.61	2.05
Gruesome	0.65	0.84	3.85	0.84	6.38	2.53
Dreadful*	0.65	0.88	3.85	0.88	7.13	1.96
Terrible	0.57	0.85	3.93	0.85	6.19	2.14
Repulsive*	0.46	0.66	4.04	0.66	7.02	2.19
Worthless*	0.44	0.63	4.06	0.63	5.53	2.37
Hateful	0.40	0.67	4.10	0.67	7.25	1.91

Note. * = adjective included in Study 2.