Counting Every Thought: Implicit Measures of Cognitive Responses to Advertising

Yanliu Huang

J. Wesley Hutchinson*

University of Pennsylvania

April, 2006

* Yanliu Huang (email: yanliuh@wharton.upenn.edu) is a doctoral candidate in marketing and J. Wesley Hutchinson (email: jwhutch@wharton.upenn.edu) is Stephen J. Heyman Professor and professor of marketing in the Wharton School, University of Pennsylvania, Suite 700 Jon M. Huntsman Hall, Philadelphia, PA 19104-6371. The authors thank the Wharton Behavioral Laboratory for support in data collection, and Dawn Iacobucci and Patti Williams for comments on prior versions of this manuscript.

Counting Every Thought: Indirect Measures of Cognitive Responses to Advertising

ABSTRACT

Prior methods of assessing consumers' cognitive responses to advertisements have focused on thought-listing techniques, and valence-weighted sums thought counts are often reliable indicators of attitudes towards products and brands. Our research explores new implicit measures of cognitive responses that focus on detecting the effects of specific thoughts. In four experiments, we demonstrate that consumers' thoughts about persuasive messages can be assessed by both an explicit memory task that involves recognition of a thought that might have occurred during an ad exposure and an implicit memory task that involves verification of a current belief and makes no reference to the previous ad exposure. Both tasks provide a way to detect the occurrence of a specific thought through its effects on memory processes. We also show that performance on these tasks (i.e., jointly observed responses, reaction times, and confidence ratings) can be modeled as Poisson counting processes and estimated model parameters are conceptually and statistically superior to any one of the three performance measures considered in isolation. Finally, our third and fourth experiments replicate these results and illustrate the effectiveness of these new implicit measures in predicting consumers' product attitudes.

Persuasive communication, as a component of advertising effectiveness, has been an ongoing topic of consumer research for over 40 years (dating back to at least Krugman 1965). Most theories posit that the thoughts elicited by persuasive messages, which are typically called *cognitive responses*, are important determinants of attitude change (e.g., Chattopadhyay and Alba 1988; Greenwald 1968; Petty and Cacioppo 1986; Shavitt and Brock 1986). In broad stroke, these theories maintain that as people process the information in advertisements, they react to it mentally and the valence (positive, negative, or neutral) of these cognitive responses contribute directly to attitude change. Developing valid measures of cognitive responses is important both to test these theories and to predict attitude change.

The traditional method of measuring cognitive responses to an advertisement is the thought-listing technique (e.g., Cacioppo and Petty 1981; Ericcson and Simon 1980; Wright 1973) in which an individual is asked to list everything coming to his or her mind during ad exposure. These thoughts are either listed concurrently as people view the ad (e.g., Brinol, Petty, and Tormala 2004; Shavitt and Brock 1986) or recalled retrospectively after some delay (e.g., Chattopadhyay and Alba 1988). Specific responses have not been used to predict attitudes. Rather, the cognitive responses are usually coded for valence and summed or averaged to form measure of net affective response. This measure is a reliable predictor of attitude (e.g., Petty and Cacioppo 1979; Wright 1973). Recently, the confidence expressed by people in the validity of their cognitive responses has been shown to also be a reliable predictor of attitude (Brinol, Petty, and Tormala 2004; Petty, Brinol, and Tormala 2002).

In this paper, we investigate the key assumptions (1) that specific cognitive responses

occur during the processing of persuasive messages and (2) that each such response exerts some effect on subsequent attitudes. Despite their success in predicting attitudes, thought-listing techniques are necessarily introspective. As such, they are suspect because retrospective methods rely on memory processes that are potentially errorful and biased, and concurrent methods have potential problems of reactivity that could distort the measurement of cognitive responses and the processing of ad information itself. In the extreme, cognitive responses might not occur at all, but thought-listing might be correlated with some other process that contributes to attitude formation and change (e.g., Nisbett and Wilson 1977). For example, thought-listing relies on intentional retrieval to report internal cognitive processing, and this intentional retrieval has the potential to introduce bias either through an unwillingness (e.g., when thoughts are socially inappropriate) or an inability (e.g., forgetting) to report thoughts that did occur or through introducing new thoughts that were not present during ad exposure (Cacioppo, von Hippel, and Ernst 1997). To address these concerns, we desire methods that have a clear causal relationship to the occurrence of specific cognitive responses and are less susceptible to the contaminating biases of introspection.

The first method we consider is thought recognition (which is described by Cacioppo et al. 1997, but we cannot find any published research using the method). In this paradigm (as we implement it), a common sentence frame is used. An example is "When viewing advertisements, people often have a variety of thoughts and reactions. Regardless of what you believe now, when viewing the magazine ads earlier in this experiment, did you ever think to yourself, 'The Fiat SUV is <u>fast</u>'?" The final sentence varies across trials and the underlined word occurs after a

delay long enough to assure comprehension of the sentence frame and brand before the target attribute is presented. The target attribute is suggested or implied by the target ad but is not explicitly mentioned in the ad. Subjects respond "Yes" or "No" and then provide a confidence rating for their response. This paradigm is a variation of the classic sentence verification task (e.g., Rosch et al. 1976; Smith, Shoben, and Rips 1974). Thought recognition differs from sentence verification mainly in being an episodic rather than a semantic memory task and in having no objectively correct response.

Thought recognition requires some introspection because the respondent is asked to examine a verbal representation of a cognitive response and affirm or deny that it occurred during ad exposure. Nonetheless, this measure offers several potential advantages over thought listing. First, recognition is generally a more sensitive measure than recall (e.g., Gardiner 1988; Johnston, Dark, and Jacoby, 1985). Second, the required information retrieval processes for recognition are simpler and less susceptible to constructive biases than recall (e.g., Alba and Hasher 1983; Novemsky and Ratner 2003). Finally, reaction times and confidence ratings for recognition tasks have been extensively studied for external stimuli, such as lists of words, pictures, and sentences (e.g., Berger and Mitchell 1989; Greenwald, McGhee, and Schwartz, 1998; Van Zandt and Maldonado-Molina, 2004). In theory, the recognition of internal, mental events should yield similar results. Reaction times have the added benefit of being unobtrusive (subjects need not know that time is being measured) and resistant to conscious control and the biases created by such control (e.g., Fazio, Powell, and Williams 1989; Fazio et al. 1986; Greenwald, McGhee, and Schwartz, 1998). Thus, we consider reaction time to be *implicit*, but

consider the response itself, valence, and subsequent confidence rating to be *explicit* because of their direct reference to the occurrence of a specific cognitive response. Our reasons for this usage are similar to those of Fazio and Olson (2003) and Whittlesea and Wright (1997) regarding implicit memory and attitude measures. We mean the term implicit to cover measures that are either unavailable to awareness or not easily controlled consciously. We do not mean to imply there are separate memory or evaluation systems.

The second paradigm we consider is belief verification. As for thought recognition, a common sentence frame is used. An example is "The Fiat SUV is <u>fast</u>." The underlined word occurs after a delay long enough to assure comprehension of the sentence frame and brand before the target attribute is presented. Subjects respond "Agree" or "Disagree." This too is a variation of the classic sentence verification task, differing only in that there is no objectively correct response. Belief verification should be even less susceptible to contamination than thought recognition because the task makes no reference to the advertisement. Thus, all measures derived from this task are implicit.

To summarize, there are three measurement paradigms that are potentially appropriate for detecting the occurrence of cognitive responses: thought listing, thought recognition, and belief verification. From each of these tasks, four specific measures are possible: the response itself, the valence of that response, the confidence expressed in the validity of that response, and the reaction time taken to respond. Table 1 summarizes the current literatures in consumer research and psychology on these methods and measures. Our goal is to empirically assess the value of thought recognition and belief verification as measures of cognitive responses to advertising.

Insert Table 1 and Figure 1 about here

EXPERIMENTAL HYPOTHESES

All of the experiments reported here manipulate the likelihood that an advertisement will produce a specific cognitive response. In particular, pre-tests identified four magazine advertisements that yielded a common response for all or almost all subjects in a retrospective thought listing task (i.e., high thought likelihood ads, which we also call target ads). Control ads (i.e., low thought likelihood ads) were constructed by removing all verbal and visual cues that suggested the target attribute. These eight ads are shown in Figure 1.

From the literatures on cognitive responses, memory, and confidence discussed thus far, we conclude that the thoughts that people have when reacting to advertisements should leave a measurable memory trace. Moreover, the tasks that should reveal that memory trace include thought recognition and belief verification for sentences that express those thoughts. In these tasks, the memory trace should reveal itself in a variety of ways, and these expected effects of memory traces form the basis of our experimental hypotheses.

First, target ads should yield more positive responses than control ads (H1, see Exhibit 1). This is essentially a manipulation check. However, this hypothesis can fail even though thought likelihood has been successfully manipulated because there are floor or ceiling effects, because there are normative or self-expressive reasons to give positive or negative responses, or because the memory system in which the trace exists is not available to conscious awareness.

Insert Exhibit 1 about here

Second, target ads should yield faster positive responses than control ads (H2, see Exhibit 1). For both tasks, the cognitive response is a form of elaboration that should enhance the accessibility of the information required to make a positive response. The mechanism is similar to that postulated for many standard priming and sentence verification tasks (e.g., Burroughs and Feinberg 1987; Shrum, Wyer, and O'Guinn' 1998; Williams, Fitzsimons, and Block, 2004) as well as for attitude judgments (e.g., Berger and Mitchell 1989; Fazio, Powell, and Williams 1989; Fazio, et al. 1986). We note, however, that the logic behind this hypothesis for positive responses offers little that helps predict latencies for negative responses. This is because the evidence used for negative responses in these tasks is neither obvious based on common sense nor well-known in the related literatures. For example, it might be that the cognitive response, in addition to increasing the accessibility of positive evidence, also decreases the accessibility of negative evidence and negative responses should be slowed. However, it could also be that negative responses are given only when an insufficient amount of positive evidence has been retrieved from memory after some internal "deadline." In this case, there would be no difference in reaction times between target and control ads. In fact, because we construct control ads by removing certain components of target ads, it could be that our target ads create more overall elaboration and therefore reduced latencies for both positive and negative responses. We will return to this issue we describe a specific model of decision making for these tasks, but at this

point we are agnostic about negative responses.

Finally, target ads should yield higher confidence ratings for positive responses than control ads (H3, see Exhibit 1). Confidence should increase because more positive evidence is retrieved from memory. Also, the rapid response itself may be interpreted as increased certainty and slow responses as uncertainty in much the same way that feelings of knowledge and feelings of learning increase confidence (e.g., Hawkins and Hoch 1992; Sanbonmatsu, Kardes, and Sansone 1991). As for reaction times, it is not clear what should be predicted for negative responses.

A PROCESS MODEL OF RESPONSE, REACTION TIME, AND CONFIDENCE

Our hypotheses involve three related dependent measures and draw upon theory from several substantive domains. However, the interrelationships among the three measures are not clear and, as noted, it is difficult to make clear predictions for negative responses. Fortunately, detailed statistical models for choice, reaction time, and confidence have been developed for simple perceptual judgments and for memory tasks. One of these, the Poisson counting model, has emerged as particularly useful for unifying the three dependent measures.[1] Thus, we will use this model to estimate a single latent measure of *memory trace strength* from joint observations of the three objectively measured variables.

---

[1] There are extensive literatures on models that jointly account for choices and reaction times (e.g., Townsend and Ashby 1983) and for choices and confidence ratings (e.g., Smith and Vickers 1988); however, joint models of all three variables are relatively few and relatively recent (e.g., Van Zandt 2000; Van Zandt and Maldonado-Molina 2004). The Poisson model is useful mainly because of it's mathematical tractability, intuitive mapping into the three variables, and empirical success in accounting for experimental results. However, this is not to say it is universally accepted as the best model and has some clear inadequacies for certain experimental paradigms and certain aspects of reaction data (e.g., see Ratliff and Smith 2005). Nevertheless, it, and the conceptually similar accumulator model of Vickers, have been particularly successful in relating confidence to choice and reaction time and have avoided critical problems that have emerged for random walk models (see van Zandt 2000).

Counting models of choice assume that memory processes generate evidence (of which people may or may not be consciously aware) in favor of each of the choice alternatives. For thought recognition and belief verification tasks, the choice alternatives are positive or negative responses. Each accumulation process simply counts the number of evidence events that favor its response. There is a criterion level of evidence for each counter and whichever counter reaches its criterion first determines the response. Essentially, the process is race.

Insert Figure 2 about here

The details of the counting model are illustrated in Figure 2 (and developed mathematically in Appendix A). In this example, the criterion for a positive response, K_p, is 5, and the criterion for a negative response, K_n, is 6. Evidence is assumed to be discrete with interarrival times that are independently and exponentially distributed. Thus, each counter is a Poisson process. The rate of evidence accumulation is v_p for the positive counter and v_n for the negative counter. The decision process superimposes the two counting processes. In the example, the positive counter reached 5 at $t = 2.7$, at which time the negative counter was at 4. As a result, the response was positive and the reaction time was 2.7. Confidence is naturally represented by the "balance of evidence" at the time of the response (e.g., Smith and Vickers 1988). In this case, there were 5 evidence events in favor of a positive response and 4 in favor of a negative response. Thus, we expect confidence to be higher than when negative evidence is at 5, but lower than when it is at 3, 2, 1, or 0 (and negative evidence cannot be 6 or more for a

positive response). That is, confidence is assumed to be inversely related to the amount of evidence supporting the opposite response. Note that, for a given response, evidence in favor of that response will be constant by definition (i.e., K_{p} or K_{n}) and can have no effect on confidence.

To account for the data in our experiments we assume that, for a given product, (1) the criteria (i.e., K_{p} and K_{n}) are the same for both target and control ads and (2) the rates at which evidence accrues (i.e., $v_{\mathrm{p}T}$, $v_{\mathrm{n}T}$, $v_{\mathrm{p}C}$, and $v_{\mathrm{n}C}$) depend on both the type of evidence and the type of ad. These model parameters (with a few technical assumptions, see Appendix A) jointly predict the likelihood of any multivariate observation of response, reaction time, and confidence rating. Thus, standard maximum likelihood methods can be used to estimate the parameters and test hypotheses about them.

Conceptually, our earlier qualitative predictions all derived from the central hypothesis that cognitive responses leave a memory trace that affects all three measures (plus the assumption that cognitive responses are more likely for target ads than control ads). The effects of these memory traces are represented in the positive rate parameters of the counting model. Thus, $v_{\mathrm{p}T}$ should be greater than $v_{\mathrm{p}C}$ (H4 in Exhibit 1), and this hypothesis is more directly related to theoretical mechanisms (i.e., memory trace strength) than are hypotheses about the three observed dependent measures (i.e., H1 - H3). As discussed earlier, predictions about negative responses are not obvious based on prior research. However, we cannot neglect them. Of particular concern, if the target ad simply elicits more elaboration than the control ad, then subsequent processing of all target ad information might be facilitated and H1 - H4 might hold because of this whether the specific cognitive response that we hypothesize occurred or not.

Because of the possibility of such a main effect, we hypothesize that any facilitation that occurs should be greater for positive responses than for negative responses. To be precise, we hypothesize that $v_{pT} - v_{pC} > v_{nT} - v_{nC}$ (H5 in Exhibit 1). Finally, the Poisson counting model predicts a negative correlation between reaction time and confidence (see Appendix A). This is also implied by H2 and H3, and we, in fact, observed a significant negative correlation between reaction time and confidence in every experiment.

In summary, estimating the counting model allows us to test our central hypothesis more directly and completely than when each dependent measure is considered in isolation. It also provides a unifying explanation for our predictions about those measures that allows us to rule out alternative explanations when those predictions are confirmed by data. Importantly, the model also provides a potential explanation when the predictions are not confirmed. In particular, H4 can hold, but H1, H2, or H3 can fail because of the values of v_{nT} and v_{nC} (see Appendix A). Thus, H4 and H5 are, in theory, better tests of our central hypothesis than are H1 - H3. Pragmatically, they are somewhat harder to test empirically because they require estimation of a specialized statistical model. As will become evident, however, this pragmatic cost is not prohibitive and, for our data at least, the results are more statistically powerful than the results of testing each dependent measure separately (i.e., testing H1 - H3).

EXPERIMENT 1

Experiment 1 was designed to test H1 - H5 for the thought recognition task.

Method

Materials and design. As discussed earlier, four target ads that were likely to elicit target

attributes as cognitive responses were selected on the basis of pretest results. These four ads are shown in Figure 1. A control ad was created for each of these by eliminating factors that could elicit target thoughts; however, brand names and product pictures were otherwise the same (see Figure 1). Each participant was presented with one ad for each product of which two were target ads and two were control ads. A fractional design was used to create orthogonal within subject factors for thought likelihood (high/target vs. low/control) and product (SUV, chipnuts, television, or lotion), plus a between subjects factor that counterbalanced stimulus order (i.e., a random order and its reversal). As noted, products were nested within price vs. non-price attributes; however, this factor is not of theoretical interest and did not produce significant effects, so it is not discussed further. This design required eight separate groups, and subjects were randomly assigned to groups.

Participants and Procedure. The participants were 150 undergraduates from an East Coast university who participated in fulfillment of a course requirement.

All materials and instructions were presented on a computer located in individual cubicles. First, demographic information was collected. Participants were then told that we were conducting marketing research concerning the effects of advertising and that they would be presented with four advertisements one by one. They were asked to view these advertisements as if they were reading a magazine and thinking that they might purchase these products. Four ads were then presented sequentially at a rate of 20 seconds per ad.

After the ad presentations, participants were given instructions about the thought recognition task. They then responded to eight specific questions: four practice questions,

followed by four experimental questions. Before each question, a dialog box appeared reminding participants to put their left index finger on the "No" key and their right index finger on the "Yes" key. They then pressed the space bar with either their right or left thumb to make a question appear. A laminated paperboard covered the keyboard except for the keys to be used in the task, which were labeled according to their response meaning. After pressing the space bar, they would first see a question with the last word missing (e.g., the first practice question was "When you entered the behavioral lab, was the experimenter ___?"), and they were instructed to read this question and wait for the last word to appear. There was a 10 second interval before the last word appeared (which was "seated" in this example). After the final word appeared, they either pressed the "No" or the "Yes" key. Following this response, a question mark appeared in the center of the screen, prompting participants to enter a confidence rating. If they responded "Yes" to the question, then they were prompted to press one of three keys labeled "Certain Yes," "Probably Yes" or "Maybe Yes" with their right index finger on the computer keyboard. If they responded "No" to the question, then they were prompted to press one of three keys labeled "Certain No," "Probably No" or "Maybe No" with their left index finger. This method is similar to that of Van Zandt and Maldonado-Molina (2004). Participants were instructed to answer each question as quickly as they could while making as few mistakes as possible, but they were not explicitly told that their responses were being timed.

Practice questions asked some general questions about the experimenter or the lab. The first three were irrelevant to the experimental task. However, the format of the final practice question was similar to the experimental questions: "When viewing advertisements, people often

have a variety of thoughts and reactions. Regardless of what you believe now, when viewing television commercials, do you ever think to yourself, 'That product is cool'?" An example of an experimental question is, "When viewing advertisements, people often have a variety of thoughts and reactions. Regardless of what you believe now, when viewing the magazine ads earlier in this experiment, did you ever think to yourself, 'The Fiat SUV is fast'?" Again, it is important to note that the target words (i.e., "fast," "spicy," "expensive," and "cheap") were not explicitly present in the ads. Participants made thought recognition judgments for all four products in a sequence consistent with the initial ad presentation order.

Dependent Measures

Response. The frequencies for each type of each type of thought recognition response (i.e., Yes or No) were explicit measures of the occurrence of the target cognitive response and, as such, acted as a manipulation check for thought likelihood.

Reaction Time (RT). RT was defined as the time between the appearance of the target word and the recognition judgment (i.e., when participants pressed the "Yes" or "No" key) and provided an imlicit measure of the occurrence of the target cognitive response.

Confidence Rating. After choosing either the positive or the negative response for each question, participants indicated their confidence using the three-point scale described earlier. This provided a second explicit measure of the occurrence of the target cognitive response.

Results

Response. Response frequencies are reported in Table 2. This binary variable was submitted to a logistic regression with thought likelihood (high/target vs. low/control) as the

predictor. Consistent with H1, more positive responses were observed for target ads (73%) than for control ads (45%). The effect of thought likelihood on response was statistically significant ($\chi^2(1) = 44.10$, $p < .0001$). H1 was also supported when each product was analyzed separately ($\chi^2(1) = 18.41$, $p < .0001$, for lotion, $\chi^2(1) = 27.47$, $p < .0001$, for SUV, $\chi^2(1) = 4.36$, $p = .03$, for TV, and $\chi^2(1) = 3.05$, $p = .08$, for chipnuts).

Insert Table 2 Here

Reaction Time. Least-square mean RTs are reported in Table 2.[2] Prior to statistical analyses, we examined the distributions of RTs and located small proportions of extremely fast and extremely slow responses. Given that these outliers are most likely due to factors not of theoretical interest, we excluded observations below 250 ms and above 7500 ms (38 out of 1200 RTs; see Greenwald, McGhee, and Schwartz 1998).

H2 and H3 pertain only to positive responses, so negative responses must be treated as missing. This precludes using traditional repeated measures ANOVA because of its reliance on the method of moments. Maximum likelihood methods, however, do not exhibit this missing data problem (for discussions of this issue see McCulloch 2005 and Wolfinger and Chang 1998). PROC MIXED in the SAS suite of statistical software was used. This method explicitly models the error covariance structure among repeated measures and, therefore, requires a choice among

[2] Throughout we report least-square means because analyzing only positive responses introduces non-orthogonalities in cell sizes and because we include Base RT as a covariate in most analyses. The qualitative pattern of the results is the same when sample means are computed.

possible structures (i.e., compound symmetry, Huynh-Feldt, and unstructured). As recommended by Wolfinger and Chang (1998), all three models were estimated and compared. A more complex model was used only if both a significant likelihood ratio test and a lower AIC value were obtained. Because of the fractional factorial design a fully crossed model is not estimable (by any method). Therefore, the estimated model included main effects of thought likelihood, product, and position, plus the interactions of thought likelihood with product and with position. The model exhausts the available degrees of freedom from experimental conditions and provides appropriate tests of H2 and its possible moderation by product or position. An individual level variable, Base RT, was computed as the average RT for the four practice questions and was included in the model as a covariate to account for individual differences in overall speed.

The model was estimated using the three error structures and criteria discussed earlier, and the unstructured model was chosen as most appropriate for these data. The statistically significant effects were thought likelihood, $F(1,144) = 9.5$, $p = .0025$, product, $F(3,144) = 10.9$, $p < .0001$, and position, $F(3,144) = 4.2$, $p = .007$. Base RT was also significant, $F(1,144) = 80.8$, $p < .0001$. The interaction of thought likelihood with product approached significance, $F(3,144) = 2.3$, $p = .08$, reflecting variation in the size, but not the direction, of the effect of thought likelihood. The effect size ranged from 27 msec for chipnuts ads to 790 msec for SUV ads with an overall average of 398 msec (see Table 2). Overall, H2 was supported.

For completeness, the RTs for negative responses were analyzed using the models, methods, and criteria discussed earlier. The compound symmetry model was chosen as most appropriate for these data. Only Base RT was statistically significant, $F(1,135) = 57$, $p < .0001$.

Confidence Ratings. Least-square mean confidence ratings are reported in Table 2. The same model, methods, and selection criteria as were used for RTs were used to analyze confidence ratings. The compound symmetry model was chosen as most appropriate for these data. The statistically significant effects were thought likelihood, $F(1,90) = 17.2$, $p < .0001$, and product, $F(3,187) = 12.7$, $p < .0001$. The interaction of thought likelihood with product approached significance, $F(3,187) = 2.5$, $p = .06$, reflecting variation in the size, but not the direction, of the effect of thought likelihood. Overall, H3 was supported.

For completeness, confidence ratings for negative responses were also analyzed using the models, methods, and criteria discussed earlier, and the compound symmetry model was chosen as most appropriate for these data. No effects were statistically significant.

Counting model. The counting model introduced earlier was estimated separately for each product (i.e., target/control pairs of ads) by maximizing the likelihoods of the jointly observed responses, RTs, and confidence ratings. Because of the strong effects of Base RT, RTs were normalized by regressing raw RTs onto Base RT and then adding a constant to the residuals to insure positivity. This normalization removed variance due to heterogeneity, which was not explicitly modeled in the analysis. The estimation method is described in Appendix A.

Insert Table 3 here

Table 3 reports estimated model parameters, goodness-of-fit indices, and statistical tests for H4 and H5. Goodness-of-fit was fairly high for all products with the possible exception of

lotion. As discussed earlier, our central hypothesis is that cognitive responses leave a memory trace that affects all three dependent measures. Given the assumption that cognitive responses are more likely for target ads than control ads, this implies that v_{pT} should be greater than v_{pC} (H4 in Exhibit 1). Table 3 reports $D_1 = v_{pT} - v_{pC}$, and t_1 provides a statistical test of $D_1 > 0$. The value of t_1 was at least 2 (i.e., $p < .05$) for all four products, supporting H4. Table 3 also reports $D_2 = v_{nT} - v_{nC}$, and the values of t_2 are less than -2 for all four products which, when combined with the results for positive responses, strongly supports H5. Although we did not predict this a priori, the fact that v_{nT} is less than v_{nC} suggests that the memory trace left by the cognitive response inhibits negative responses and well as facilitating positive response. The mechanism of this inhibit might be similar to that for part-list cuing effects in which items recalled early inhibit the recall of other items (e.g., Alba and Chattopadhyay 1985).

EXPERIMENT 2

The pattern of results observed in Experiment 1 was supportive of all of our hypotheses regarding thought recognition as a measure of cognitive response (i.e., H1 - H5). Compared to control ads, target ads yielded more positive responses that were given more quickly and with greater confidence. In a thought recognition task, only RT is an implicit measure. The response and rated confidence in the response are potentially susceptible to contamination because they explicitly refer to the possible occurrence of the cognitive response. Thus, it is theoretically and pragmatically important to develop a more completely implicit measure of cognitive responses. Experiment 2 was conducted to examine one such method using the belief verification task.

Method

All experimental materials, design, and procedure were identical to that used in experiment 1, except that participants were required to perform a series of belief verification tasks rather than thought recognition tasks after seeing four ads.

Participants. One hundred and fifty-one students from an East Coast university received $10 monetary compensation for participating in the experiment.

Dependent Measures

The dependent measures were the same as those of experiment 1 (i.e., response frequency, RT, and confidence), except the task was belief verification. All three measures can be regarded as implicit insofar as the task does not explicitly refer to the prior advertisements.

Results

Response. Response frequencies are reported in Table 2. This binary variable was submitted to a logistic regression with thought likelihood (High/target vs. Low/control) as the predictor. Consistent with H1, more positive responses were observed for target ads (88%) than for control ads (75%). The effect of thought likelihood on response was statistically significant ($\chi^2(1) = 16.0$, $p < .0001$). When this analysis was conducted separately for each product, all were directionally consistent with H1 and two were statically significant ($\chi^2(1) = 3.95$, $p = .04$, for lotion, and $\chi^2(1) = 19.66$, $p < .0001$, for SUV, $\chi^2(1) = 1.87$ for TV, and $\chi^2(1) = 1.05$ for chipnuts). Overall, H1 was supported; however, positive responses were more likely in general for belief verification than for thought recognition, and there appears to have been a ceiling effect for the TV and chipnuts ads.

Reaction Time. Mean RTs are reported in Table 2 and Figure 2. As for experiment 1, RTs below 250 ms and above 7500 ms were excluded from analysis (33 out of 1208 RTs). RTs for positive responses were analyzed using the models, methods, and criteria described earlier, and compound symmetry was chosen as most appropriate for these data. The statistically significant effects were thought likelihood, $F(1,132) = 30.6$, $p < .0001$, product, $F(3,322) = 13.0$, $p < .0001$, and position, $F(3,322) = 5.1$, $p = .002$. Base RT was also significant, $F(1,322) = 98.8$, $p < .0001$. The interaction of thought likelihood with product was also significant, $F(3,322) = 2.9$, $p = .04$, reflecting variation in the size, but not the direction, of the effect of thought likelihood. The effect size ranged from 170 msec for chipnuts ads to 955 msec for SUV ads with an overall average of 495 msec (see Table 2). Overall, H2 was supported and the effect was larger and stronger than for thought recognition in experiment 1. We note that the chipnuts ads and the SUV ads also had the smallest and largest effects (respectively) in experiment 1. This suggests some stability in the moderating effect of product.

RTs for negative responses were also analyzed using the models, methods, and criteria described earlier, and compound symmetry was chosen as most appropriate for these data. The only statistically significant effect was Base RT, $F(1,77) = 16.6$, $p = .0001$.

Confidence Ratings. Least-square mean confidence ratings are reported in Table 2. Ratings for positive responses were analyzed using the models, methods, and criteria described earlier, and compound symmetry was chosen as most appropriate for these data. The statistically significant effects were thought likelihood, $F(1,132) = 64.7$, $p < .0001$, product, $F(3,322) = 20.4$, $p < .0001$, and position, $F(3,322) = 2.8$, $p < .04$. The interaction of thought likelihood with

product was also significant, $F(3,322) = 2.7$, $p = .04$, reflecting variation in the size, but not the direction, of the effect of thought likelihood. Overall, H3 was supported.

Confidence ratings for negative responses were also analyzed using the models, methods, and criteria discussed earlier, and compound symmetry was chosen as most appropriate for these data. The only statistically significant effect was Base RT, $F(1,77) = 8.2$, $p = .005$.

Counting model. The counting model was estimated separately for each product using the same method as in experiment 1. Table 3 reports estimated model parameters, goodness-of-fit indices, and statistical tests for the best fitting models. Goodness-of-fit was fairly high for all products with the possible exception of chipnuts. As in experiment 1, the statistical test of H4, t_1, was at least 2 (i.e., $p < .05$) for all four products. Also, t_2 was less than -2 for all products except TV. This supports H5 and replicates the inhibitory effect of target thoughts on the accumulation of negative evidence that was observed in experiment 1.

Discussion

The results for belief verification in experiment 2 closely replicated the results for thought recognition that were observed in experiment 1. Thus, we have shown that people's thoughts about an advertisement can be measured implicitly without making reference to a previous ad exposure. In fact, the effects appear to be stronger for belief verification than for thought recognition. Given that we have demonstrated the effectiveness of two new techniques for detecting specific cognitive responses, it is important to examine the ability of these measures to predict attitudes. Experiment 3 was designed to achieve this purpose.

EXPERIMENT 3

As discussed in the introduction, our goal is to empirically test two assumptions made by most current models of persuasion and advertising effectiveness: (1) specific cognitive responses occur during the processing of persuasive messages, and (2) each such response exerts some effect on subsequent attitudes. Experiments 1 and 2 have provided strong evidence for the first assumption by showing that the occurrence of a specific target thought during ad exposure results in measurable effects on response frequencies, RTs, and confidence ratings for both thought recognition and belief verification tasks. In experiment 3, we tested the second assumption by using these measures as predictors of attitude.

Prior research has almost exclusively used thought-listing measures and measures derived from listed thoughts, such as valence, confidence in thought validity, and classification as an interpretation, abstraction, or global evaluation (e.g., Brinol, Petty, and Tormala 2004; Chattopadhyay and Alba 1988; Ericson and Simon 1980; Petty, Brinol, and Tormala 2002; Petty and Cacioppo 1979; Shavitt and Brock 1986). Perhaps more importantly, only overall indices computed by aggregating the derived measures over thoughts for each subject have been used as predictors of attitude. These aggregate indices have the advantage of accommodating heterogeneity across individuals in the specific content of cognitive responses. However, they strip each thought of it's unique content. This increases the plausibility of alternative explanations of the observed correlation between valenced elaboration (or confidence) and attitude. For example, some third factor, such as non-conscious reactions to the ad, might jointly cause both changes in attitude and the recall of attitude consistent thoughts, even though those

thoughts had no causal impact on attitude. This explanation is possible, but much less plausible, for specific thoughts that are defined a priori and based on specific content in the ad. More pragmatically, such content specific measures potentially provide guidance for improving an ad and are, therefore, potentially useful for copy testing purposes (cf. Wansink et al. 1994).

Prior research and the results of experiments 1 and 2 suggest the following experimental hypotheses for the effects of measures of specific thoughts when included as independent variables in regression models of attitude (see Exhibit 1). First, there should be a main effect of valence on attitude (H6). Second, there should be a main effect of RT on attitude (H7a) that is moderated by an interaction with valence (H7b). Finally, there should be a main effect of confidence on attitude (H8a) that is moderated by an interaction with valence (H8b). These hypotheses are reasonable only when a positive response is given. As for RTs and confidence, current models of persuasion have little to say about thoughts that do not occur during exposure to a message. It might be argued that beliefs that are denied should have effects opposite to those of beliefs that are verified. However, it is difficult to know whether a negative response was due to a belief in the opposite of the target thought (e.g., "The Fiat SUV is slow.") or to a high level of uncertainty about the validity of the claim. Thus, we simply include an omnibus hypothesis for belief verification, but not for thought recognition, that the effects obtained when the response is positive should be reversed when the response is negative (H9).

Method

All the materials, design and procedure adopted in the present study were the same as in experiment 1 and 2 with the following changes. First, participants' product evaluations were

collected right after viewing all four ads and prior to the cognitive response tasks. Putting attitude measures ahead of cognitive response measures makes it less possible that participants' product evaluations would be contaminated by how they respond in the cognitive response tasks. Also, the relatively long instructions on how to perform the cognitive response tasks serve as a good distracter to lessen the possible effect of attitude measures on participants' performance. Recall that the cognitive response tasks are designed to measure the changes in memory trace strength that occur during ad exposure and before product evaluation. The second methodological change was that several questions were asked at the end of the experiment as a means to assess the valence of the target product attributes.

Participants. One hundred and forty-four students from an East Coast university received $10 monetary compensation for participating in the experiment. Participants were randomly assigned to either the thought recognition task (N=83, referred to as experiment 3a) or the belief verification task (N=61, referred to as experiment 3b).

Dependent Measures

The dependent measures included those used in experiments 1 and 2 (i.e., response frequency, RT, and confidence). Additionally, attitudes toward each of the four products were assessed using four seven-point semantic differential scales with the following end points: negative-positive, bad-good, favorable-unfavorable, and dislike-like. Valence was measured by a seven-point Likert scale rating for each target thought using the sentence frame, "Assume that <*target thought*> is true, do you think it will increase the attractiveness of this product?" where the target thoughts were those used in the thought recognition and belief verification tasks.

Results

The results of experiment 3 replicated those of experiments 1 and 2 (see Table 2). The data were analyzed as in the earlier experiments. Consistent with H1, the effect of thought likelihood on response was statistically significant for thought recognition ($\chi^2(1) = 16.2$, $p < .0001$) and was marginally significant for belief verification ($\chi^2(1) = 2.74$, $p = .09$). For positive responses in the thought recognition task, the statistically significant effects on reaction time were thought likelihood, $F(1,50) = 3.84$, $p = .056$, product, $F(3,95) = 6.57$, $p = .0004$, and Base RT, $F(1,80) = 28.2$, $p < .0001$. The interaction of thought likelihood with position was also significant, $F(3,95) = 2.74$, $p = .048$, reflecting a reversal of the effect of thought likelihood in one of the four positions. For positive responses in the belief verification task, the statistically significant effects on reaction time were thought likelihood, $F(1,54) = 18.7$, $p < .0001$, and product, $F(3,117) = 5.48$, $p = .002$. Base RT was also significant, $F(1,58) = 52.0$, $p < .0001$. Overall, H2 was supported for both thought recognition and belief verification. We also note that the effect of thought likelihood was larger and more reliable for belief verification than for thought recognition, also replicating the results of Experiments 1 and 2. For thought recognition, the statistically significant effects on confidence ratings were thought likelihood, $F(1,50) = 7.82$, $p = .007$, and product, $F(3,95) = 6.48$, $p = .0005$. For belief verification, the statistically significant effects on confidence ratings were also thought likelihood, $F(1,54) = 28.2$, $p < .0001$, and product, $F(3,117) = 12.4$, $p < .0001$. Overall, H3 was supported.

Counting model. The counting model was estimated as before, and the results are given in Table 3. Goodness-of-fit was fairly high for all products with the possible exception of the

SUV for thought recognition. H4 was directionally supported in every case, and t_1 was greater than 2 (i.e., $p < .05$) for the same three out of four products in both data sets. Also, t_2 was less than -2 for all products for thought recognition and for chipnuts and TV for belief verification. Thus, H5 was supported and replicates the inhibitory effect of target thoughts on the accumulation of negative evidence that was observed in the earlier experiments.

Attitudes. The four attitude measures demonstrated high internal consistency (Cronbach $\alpha = .97$ for thought recognition, and Cronbach $\alpha = .96$ for belief verification), so they were averaged to form a composite attitude index.

Our approach to testing hypotheses H6 through H9 was to use a linear regression model that included main effects for all variables of interest (i.e., product, thought likelihood, response, valence, RT, and confidence), plus interactions of those variables with valence (because of H7 and H8), with response (because of H6 and H9), and with product (as a test of robustness for all effects). To simplify interpretation of the results and reduce multicollinearity, valence was a binary class variable when interacted with other variables (i.e., valence defined to be was "positive" if the rating was greater than 4 and "negative" if it was 4 or less; this was approximately a median split of the observed ratings). Observed ratings were used directly for the main effects of valence and for confidence. To remove individual differences, RTs were zero-centered for each subject prior to the analysis. As for previous analyses, maximum likelihood was used to estimate the model, including correlated error due to the repeated measures design, and the criteria described earlier were used to select the best error structure.

For thought recognition, the analysis showed that valence (and it's interactions)

significantly improved the model compared to a base model that only had effects of thought likelihood, product, and their interaction, $\chi^2(14) = 73.4$, $p < .0001$. However, adding RTs and confidence (and their interactions) only marginally improved the model, $\chi^2(14) = 23.0$, $p = .06$. Also, when each effect in the model was tested separately only product, $F(3,193) = 4.22$, $p = .006$, valence, $F(1,193) = 4.20$, $p = .042$, valence x product, $F(3,193) = 3.07$, $p = .029$, and RT x product, $F(3,193) = 2.68$, $p = .049$, were significant. Moreover, except for valence, the signs of the coefficients were generally not consistent with our hypotheses. Thus, we conclude that for thought recognition H6 was supported, but H7 - H9 were not supported.

For belief verification, the analysis showed that valence (and it's interactions) significantly improved the model compared to a base model that only had effects of thought likelihood, product, and their interaction, $\chi^2(14) = 26.4$, $p = .023$. In contrast to thought recognition, adding RT and confidence (and their interactions) significantly improved the model, $\chi^2(14) = 33.2$, $p = .003$. Also, adding each factor (and it's interactions) separately significantly improved the model ($\chi^2(5) = 14.4$, $p = .013$, for RT and $\chi^2(5) = 19.0$, $p = .002$, for confidence). When each effect in the model was tested separately the following effects were significant: product, $F(3,134) = 3.42$, $p = .020$, product x response, $F(1,134) = 6.18$, $p = .014$, product x thought likelihood, $F(3,193) = 1.34$, $p = .025$, valence, $F(1,134) = 8.00$, $p = .005$, RT x response, $F(1,134) = 5.09$, $p = .026$, RT x product, $F(3,134) = 3.63$, $p = .015$, and confidence x response x product, $F(1,134) = 10.2$, $p = .002$. These effect tests must be regarded with caution, however, because the variables and especially their interactions are not uncorrelated. Thus, we consider the nested model tests to be the most valid and interpretable. Overall, H6 - H9 were supported;

however, there were some exceptions (see subsequent discussion).

Discussion

Experiment 3 replicated the results of experiments 1 and 2 in all respects, adding support to the general conclusion that thought recognition and belief verification tasks provide reliable indicators that a specific thought occurred during the encoding of an advertisement. As in the earlier experiments, the observed effects were somewhat larger and more reliable for belief verification than for thought recognition.

We note that attitudes are determined by a wide variety of factors and our subjects are likely to have had pre-existing attitudes about the product we used. Thus, it would take very sensitive measures of both attitudes and their potential antecedents to detect differences across subjects deriving the occurrence of a single thought. From this perspective, the obtained results are impressive (especially for the belief verification task).

For the thought recognition task, experiment 3 provided support for an effect of valence on attitude (H6), but not for effects of RT (H7) or confidence (H8). Although the interaction of valence with response was not significant, the estimated coefficient was +.47 for positive responses and -.08 for negative responses, which is consistent with H9.

For the belief verification task, the nested model tests, which show effects of both reaction time and confidence, are consistent with the conclusion that a single thought during exposure to an ad can cause, or at least indicate, a change in belief strength that in turn changes attitude. Table 4 presents the estimated coefficients that are pertinent to our experimental hypotheses. Hypothesis 7a is supported because the coefficient for RT is negative when the

response and valence are positive. Consistent with Hypothesis 7b, this coefficient reverses sign when valence is negative, and consistent with H9, it reverses sign when the response is negative. However, H7 and H9 together predict that when both response and valence are negative there should be a "double reversal" or cancellation effect, and the sign of the coefficient should match that of the coefficient for positive response and valence. In contrast, this coefficient is the most strongly reversed. This suggests reasoning something like, "I don't like that attribute, and I think you are lying about it anyway," an additive rather than a multiplicative effect. Thus, this outcome is more intriguing than damaging. Also, we note that there are very few observations where both the response and the valence are negative (16 out of 231). In fact, there were only 31 negative responses, compared to 200 positive responses; thus, our test of H9 has limited statistical power. Hypothesis 8a is supported by the positive coefficient for confidence when response and valence are positive; however, H8b is not supported. As for reaction times, Hypothesis 9 is supported when valence is positive, but not when it is negative.

Insert Table 4 Here

Overall, for belief verification our hypotheses were supported most strongly when both the response and valence were positive. This is the condition in which theory makes the clearest predictions. It is also the most frequently observed condition and therefore has the greatest statistical power. As discussed earlier, it is hard to know much about the types of evidence that lead to a negative response or what effects such evidence would have on attitude. For example, a

person might have disagreed with the statement that the Fiat SUV was fast because they believed it was big and heavy. However, some people might feel that size increases safety and have a positive attitude toward the car. Other might feel that size decreases fuel efficiency and have a negative attitude. Thus, the nature of the negative evidence is unknown to us and has ambiguous effects. In experiment 4, we track multiple, specific thoughts that occur during ad exposure so that these indeterminacies can be resolved.

EXPERIMENT 4

Experiment 3 examined the effects of a single cognitive response on brand attitude. The results confirmed our hypotheses when responses and valences were positive, but the results were mixed when valence was negative and especially when the response was negative. However, these conditions were infrequently observed and lacked statistical power. We also noted that attitudes are determined by a wide variety of factors, and our subjects were likely to have had pre-existing attitudes about the advertised products. Thus, the effects of a single thought are likely to be small compared to combined effects of these factors, and it is impressive that such small effects were detected at all for belief verification.

In experiment 4, we examine the properties of response, valence, reaction time, and confidence ratings in the more complete context of a battery of statements about products that are likely to cover many, arguably most, of the types of thoughts consumer might have in response to an advertisement. This approach redresses important limitations of the single thought paradigm used in the first three experiments. However, there is an important conceptual difference between the single thought paradigm and the battery of statements paradigm. The

former is designed to detect the occurrence of a thought as a reaction to a specific persuasive communication. The latter is designed to detect differences in beliefs about a product, regardless of the source of those differences. To test the usefulness of the battery of statements paradigm we derive valence weighted indices from responses, reaction times, and confidence rating and use them to model attitudes, as is commonly done for other measures of beliefs and cognitive responses (e.g., Fishbein and Ajzen 1975; Petty and Cacioppo 1986; Toy 1982; Wyer and Albarracin 2005). Because this approach focuses more generally on beliefs (that include, but are not limited to, cognitive responses) and because our results have been stronger for belief verification than for thought recognition, we develop measures using the belief verification task.

Method

The materials, design, procedures, and dependent variables used in experiment 4 were similar to the belief verification task in Experiment 3 with a few exceptions. First, following product evaluations, participants performed 48 rather than 4 formal belief verification tasks. Specifically, in addition to the original target sentence used in the first three experiments, 5 more attributes for each product were collected from the verbal protocols generated in the pre-test, and all of these 6 attributes were framed as either negative or positive in valence, resulting in a total of 12 sentences for each product with 6 positive and the other 6 negative. Furthermore, among all 12 sentences for each product, 8 were the same across all the four products (e.g., <*Product Name*> is cool/boring.), and the other 4 were specific to each product (e.g., Fiat Punto SUV is fast/slow). In order to create short completions for all of these 48 sentences, we also varied the sentence frames between "is" (e.g., Panasonic HDTV is <u>beautiful.</u>) and "has" (e.g., Panasonic

HDTV has a clear picture.) with 24 sentences having the "is" frame and 24 having the "has" frame. Finally, as a between-subject variable, these 48 tasks were presented to participants either in a product-blocked order (with 12 sentences in each block randomized) or in a random order. All of the belief verification sentences are provided in Appendix B.

Participants. Two hundred and forty students from an East Coast university received $10 monetary compensation for participating in the experiment.

Results and Discussion

Attitude was the main dependent measure of interest in experiment 4. As in experiment 3, the four semantic differential ratings were averaged to form a composite index. Blocked vs. randomized order was an exploratory manipulation and had no statistically significant effects. Therefore, it is not discussed further.

Independent variables. The valence of each statement was defined a priori (see Appendix B). Index variables were computed at the individual level for each product. The *valence index* was computed as the number of "Agree" responses to positively valenced beliefs minus the number of "Agree" responses to negatively valenced beliefs . There was no need to compute a valence index for "Disagree" responses because this index is by definition perfectly, negatively correlated with the "Agree" index. The *RT(Agree) index* was computed as the average RT for "Agree" responses to positively valenced beliefs minus the average RT for "Agree" responses to negatively valenced beliefs . The *RT(Disagree) index*, *Confidence(Agree) index*, and *Confidence(Disagree) index* were computed analogously. The RT and Confidence indices require at least one response in each of the four response categories (i.e., "Agree" for a positive

belief, “Disagree” for a positive belief, “Agree” for a negative belief, and “Disagree” for a negative belief) because the average is undefined otherwise. Thus, all data analyses were performed on the subset of observations that satisfied this condition (i.e., 694 of 963). As for the previous experiments, only RTs greater than 250 msec and less than 7500 msec were excluded. To be conservative, if any RT was outside this range, the entire observation was excluded (recall that each observed attitude rating was associated with 12 RTs, one for each belief). This reduced the final data set to 574 observations. These indices for RT and confidence are preferred to indices computed as the sum of values for positive beliefs minus the sum of values for negative beliefs because the latter approach induces large correlations with the Valence index.

Correlational analyses. All of the belief verification indices were correlated with brand attitude, as can be seen in Figure 3. The simple correlations with attitude (N = 574) were .67 for Valence, -.27 for RT(Agree), .25 for RT(Disagree), .25 for Confidence(Agree), and -.36 for Confidence(Disagree). All correlations were in the expected direction and statistically significant ($p < .0001$), supporting H6 - H9. When the positive and negative components of the RT and Confidence indices were examined separately, each had opposite signs consistent with H7b and H8b (i.e., -.23 and .06 for RT(Agree), .06 and -.18 for RT(Disagree), .22 and -.06 for Confidence(Agree), and -.11 and .29 for Confidence(Disagree); $p < .01$ whenever $r > .10$).

Insert Figure 3 and Table 5 about here

Regression analyses. Because up to four repeated measures were available for each

subject, the same maximum likelihood methods and error covariance structure criteria that were used in the earlier experiments were used for experiment 4. Compound symmetry was found to be the best error structure for all analyses. As in experiment 3, a nested models approach was used. These models were based on three sets of variables: dummy variables for ads, dummy variables for products, and the belief verification indices described earlier.

Table 5 reports the results of estimating six models. The underlying causal model is that belief strength affects attitudes as commonly postulated (e.g., Fishbein and Ajzen 1975; Petty and Cacioppo 1986; Wyer and Albarracin 2005). Differences in belief strength can be due to both the effects of presented ads and to pre-existing individual differences. The belief verification indices described earlier measure belief strength at the individual level. The dummy variables for ads and for products are a natural baseline insofar as they represent a simple model in which all individuals have the same attitude for a given product (perhaps modified by the ad viewed prior to making their attitude judgments). The goodness-of-fit measures reported in Table 5 (i.e., -2LL, AIC, and BIC) indicate that the belief verification indices account for unique variance in brand attitude. The fact that the estimated coefficients for the ad and product dummy variables are reduced, but not eliminated by the belief verification indices, suggests that they capture some, but not all, aspects of attitude. The estimated coefficients for each belief verification index are quite stable across models. The *F* statistic is a reasonable indicator of effect size for fixed effect regressions (Rosenthal and Rosnow 1991), and it suggests that the valence index is most powerful, followed by Confidence(Agree), RT(Agree), Confidence(Disagree), and RT(Disagree), in that order.

Overall, H6, H7, and H8 are strongly supported by the results for Valence, RT(Agree), and Confidence(Agree). H9 is supported by the results for RT(Disagree) and Confidence(Disagree), but the implied effects are weak. As discussed earlier, predictions for Disagree responses are much less clear than for Agree responses.

GENERAL DISCUSSION

From one perspective, the results reported here provide strong validating evidence for the traditional thought-listing technique for measuring cognitive responses. This is because (1) the manipulation of thought likelihood was based on pretests that used thought-listing to identify advertisements for which a specific thought was frequently listed and (2) this manipulation provided reliable effects for both thought recognition and belief verification. Moreover, the estimated parameters of the counting model strongly confirmed the effects of thought likelihood. In addition to confirming that retrospectively listed thoughts occurred as cognitive responses during ad exposure, traditional valence-weighted aggregate measures of cognitive response were shown to be predictive of attitudes toward advertised products (experiments 3 and 4).

Our results also highlight several weaknesses of the traditional thought-listing paradigm, however. First, the implicit measures examined here highlight the known problems of explicit measures, such as people being unable or unwilling to accurately report their thoughts. Our stimuli were chosen to maximize the likelihood that people would be both willing and able to report thoughts. If we had not found strong evidence for the occurrence of the hypothesized thoughts, it would have been very damaging evidence against the validity of thought-listing. Therefore, an important problem for future research is to explore conditions under which people

are most likely to be unwilling or unable to directly report their thoughts and to determine whether or not the implicit measures developed here are reliable indicators of cognitive responses when thought-listing fails. Second, our results highlight the fact that traditional valence-weighted measures of cognitive response predict some, but not all, of the systematic variation in attitudes across people. The new belief verification indices developed here were shown to provide significant additional predictive power. Third, results in other areas of attitude research have shown that imlicit measures are sometimes better than explicit measures in predicting subsequent spontaneous behavior, even when they uncorrelated with attitudes (e.g., Bassili and Brown 2005; Fazio and Olson 2003). Therefore, an important problem for future research is to identify conditions under which explicit and implicit measures of cognitive response become dissociated and to develop theories to account for this dissociation. Finally, despite the success that traditional thought-listing measures have had in academic research, they have not been widely adopted in commercial advertising research, although parallels between academic and commercial measures have been noted and academic measures other than cognitive response indices have been related to in-market effectiveness (e.g., MacInnis, Rao, and Weiss 2002). One reason for this may be the lack of emphasis on specific cognitive responses in academic research (see Wansink, Ray, and Batra 1994). To guide the development of more persuasive advertising and to guide public policy about misleading advertising, measures must be able to detect the effects of specific ad elements on the cognitive responses, beliefs, attitudes, and behaviors of consumers. The approach introduced here is a first step in that direction--counting every thought.

APPENDIX A: POISSON COUNTING MODEL

The development of the Poisson counting presented here closely follows the exposition and proofs given in Townsend and Ashby (1983, pp. 36-43 and pp. 272-280). Unless otherwise stated, the results given here follow directly from Townsend and Ashby, or can be derived in a few algebraic steps.

Three random variables are observed in our data: the response **R**, which takes specific values $r \in \{\text{p, n}\}$, the time of the response, **T**, which takes specific values, t, that are positive real numbers, and a confidence rating, **C**, which takes specific values $c \in \{1, 2, 3\}$. The model assumes that decisions are based on two independent counters: $X_{\text{p}}(t)$ accumulates positive evidence events (by summation) and $X_{\text{n}}(t)$ accumulates negative evidence events. Thus, $\mathbf{X}_{\text{p}}$ and $\mathbf{X}_{\text{n}}$ are random variables that represent unobserved internal states. Evidence events are independently and exponentially distributed with rate parameters v_{p} and v_{n} for positive and negative events, respectively. The rate parameters differ across stimuli (e.g., $v_{\text{p}T}$ and $v_{\text{n}T}$ for a target ad and $v_{\text{p}C}$ and $v_{\text{n}C}$ for a controls ad). The response is determined by the first counter to reach its criterion (i.e., K_{p} for a positive response and K_{n} for a negative response). Considered separately, each counter is a Poisson process and the time to accumulate any specific number of evidence events is distributed as gamma. The complete process is a superposition of the two counting processes and is also a Poisson process (with rate parameter $v_{\text{p}} + v_{\text{n}}$).

The "balance of evidence" at t determines the confidence rating. For a positive response, $c = g(X_{\text{p}}(t) - X_{\text{n}}(t))$, and for a negative response, $c = h(X_{\text{n}}(t) - X_{\text{p}}(t))$. For a positive response,

$X_p(t)$ is K_p by definition, so, for simplicity, g is assumed to be stepwise monotonically increasing, such that c is the confidence rating whenever $c^- \leq X_n(t) \leq c^+$. For negative responses, h is a similar stepwise monotonically increasing function.

Estimation Method

The likelihood for a positive response given at time t with confidence c is

$$
\begin{aligned}
f(\mathbf{R} = \mathrm{p}, \mathbf{T} = t, \mathbf{C} = c) \;=\; & \text{Prob(Positive Evidence} = K_p - 1 \text{ and } \mathbf{T} = t) \\
& \times \text{Prob}(c^- \leq \text{Negative Evidence} \leq c^+ \text{ and } \mathbf{T} = t) \\
& \times \text{Likelihood(New Positive Evidence in the next instant of time)} \\
= \; & \frac{(v_p\, t)^{K_p - 1}\, v_p \exp(-v_p\, t)}{(K_p - 1)!} \sum_{j = c^-}^{c^+} \frac{(v_n\, t)^j \exp(-v_p\, t)}{(j)!} \; . \qquad \text{(A.1)}
\end{aligned}
$$

A simple approach was used to estimate the model from our data. First, so that a homogeneous model was plausible, all RTs were regressed onto Base RTs, and the residuals, t', were treated as the times to be modeled. The purpose of this transformation was to remove individual differences in mean RT. Maximum likelihood estimation, based on A.1 and the assumption that $t', = t_0 + t$, was implemented using the *nlm* function of the statistical programming language *R*. In particular, for each product, a model was estimated for each pair of criteria (K_p, K_n) for integer values from 3 to 6 (i.e., 16 models) that maximized the following likelihood function.

$$L[(r, t', c); (t_0, v_{pT}, v_{nT}, v_{pC}, v_{nC})] \;=\; \prod_s f(r_s, t_s, c_s) \; . \qquad \text{(A.2)}$$

The t statistics reported in Table 3 were obtained in the conventional manner from the Hessian matrix that was estimated by the *nlm* function in *R*.

To select among the 16 models, and to provide an easily interpreted index of goodness-of fit, model parameters were used to compute expected values for each of the 12 possible thought likelihood/response/confidence triples for a given product using the following formulae (here given only for positive responses).

$P(\mathbf{R} = \text{p and } \mathbf{C} = c) = P_{\text{p}c}$

$$= \sum_{j=c^-}^{c^+} \binom{K_\text{p} + j - 1}{j} \left(\frac{v_\text{n}}{v_\text{p} + v_\text{n}}\right)^j \left(\frac{v_\text{p}}{v_\text{p} + v_\text{n}}\right)^{K_\text{p}} . \tag{A.3}$$

$E(\mathbf{T} \mid \mathbf{R} = \text{p and } \mathbf{C} = c)$

$$= \frac{1}{P_{\text{p}c}} \sum_{j=c^-}^{c^+} \binom{K_\text{p} + j - 1}{j} \left(\frac{v_\text{n}}{v_\text{p} + v_\text{n}}\right)^j \left(\frac{v_\text{p}}{v_\text{p} + v_\text{n}}\right)^{K_\text{p}} \left(\frac{K_\text{p} + j}{v_\text{p} + v_\text{n}}\right) . \tag{A.4}$$

Overall variance-accounted-for (*VAF*, as reported in Table 3) was simply the R^2 obtained when observed times were regressed on to model-based expected values. The best-fitting model reported in Table 3 was selected based on overall VAF; however, the qualitative results were robust with respect to (K_p, K_n). To provide an index of "systematic" *VAF*, dummy variable regression was used to compute the best possible R^2 for models of this type, and the ratio of overall R^2 to best-possible R^2 is also reported in Table 3.

Relationship of the Model to Experimental Hypotheses

All else equal, when H4 holds, H1, H2, and H3 will also hold. However, H1 (the effect on response frequencies) could fail when H4 holds because both $v_{\text{n}T}$ and $v_{\text{n}C}$ are very small and a ceiling effect occurs for positive responses (or alternatively they are large and a floor effect occurs). H2 (the effect on reaction times) could fail when H4 holds because $v_{\text{n}C}$ is much larger

than v_{nT} (which would be consistent with H5). This result is counter-intuitive because it seems that v_{nT} and v_{nC} should not affect reaction times for positive responses. However, because the reaction times are conditional upon the response, there are indirect "competitive" effects. If v_{nC} is very large then this strong competition means that a positive response for control ads will occur only for the fastest times in the distribution of possible times for positive evidence reaching K_p. That is, the conditional mean reaction time is based on a highly truncated distribution. If v_{nT} is very small, then positive responses will be frequent, truncation will be minimal, and the reaction time will increase. In principle, these factors could overcome the effect of v_{pT} being larger than v_{pC}. For example, H2 fails when $v_{nT} = 1 < 4 = v_{nC}$, even though $v_{pT} = 2 > 1 = v_{pC}$. Finally, H3 (the effect on confidence ratings) could fail when H4 holds because of competitive effects opposite to those just described. For example, H2 fails when $v_{nT} = 1 > .25 = v_{nC}$, even though $v_{pT} = 2 > 1 = v_{pC}$. Of course, this competitive effect contradicts H5.

Some Properties of the Poisson Counting Model

Expected values for response probabilities, RT, and confidence are as follows (for simplicity g and h are assumed to be the identity function).

$$P(\mathbf{R} = \mathrm{p}) = P_{\mathrm{p}} = \sum_{j=c^-}^{c^+} \binom{K_{\mathrm{p}} + j - 1}{j} \left(\frac{v_{\mathrm{n}}}{v_{\mathrm{p}} + v_{\mathrm{n}}}\right)^{j} \left(\frac{v_{\mathrm{p}}}{v_{\mathrm{p}} + v_{\mathrm{n}}}\right)^{K_{\mathrm{p}}} . \tag{A.5}$$

$$E(\mathbf{T} \mid \mathbf{R} = \mathrm{p}) = \frac{1}{P_{\mathrm{p}}} \sum_{j=0}^{K_{\mathrm{n}} - 1} \binom{K_{\mathrm{p}} + j - 1}{j} \left(\frac{v_{\mathrm{n}}}{v_{\mathrm{p}} + v_{\mathrm{n}}}\right)^{j} \left(\frac{v_{\mathrm{p}}}{v_{\mathrm{p}} + v_{\mathrm{n}}}\right)^{K_{\mathrm{p}}} \left(\frac{K_{\mathrm{p}} + j}{v_{\mathrm{p}} + v_{\mathrm{n}}}\right) . \tag{A.6}$$

$$E(\mathbf{C} \mid \mathbf{R} = \mathrm{p}) = \frac{1}{P_{\mathrm{p}}} \sum_{j=0}^{K_{\mathrm{n}}-1} \binom{K_{\mathrm{p}}+j-1}{j} \left(\frac{v_{\mathrm{n}}}{v_{\mathrm{p}}+v_{\mathrm{n}}}\right)^{j} \left(\frac{v_{\mathrm{p}}}{v_{\mathrm{p}}+v_{\mathrm{n}}}\right)^{K_{\mathrm{p}}} (K_{\mathrm{p}} - j) \quad . \tag{A.7}$$

To see the relationship between RT and confidence, note that for positive responses the expect value of the negative counter when the positive response occurs is as follows.

$$E(\mathbf{X}_{\mathrm{n}} \mid \mathbf{R} = \mathrm{p}) = \frac{1}{P_{\mathrm{p}}} \sum_{j=0}^{K_{\mathrm{n}}-1} \binom{K_{\mathrm{p}}+j-1}{j} \left(\frac{v_{\mathrm{n}}}{v_{\mathrm{p}}+v_{\mathrm{n}}}\right)^{j} \left(\frac{v_{\mathrm{p}}}{v_{\mathrm{p}}+v_{\mathrm{n}}}\right)^{K_{\mathrm{p}}} j \quad . \tag{A.8}$$

As a result, the expected values of RT and confidence can be written as follows.

$$E(\mathbf{T} \mid \mathbf{R} = \mathrm{p}) = [K_{\mathrm{p}} + E(\mathbf{X}_{\mathrm{n}} \mid \mathbf{R} = \mathrm{p})] / (v_{\mathrm{p}} + v_{\mathrm{n}}) \tag{A.9}$$

$$E(\mathbf{C} \mid \mathbf{R} = \mathrm{p}) = K_{\mathrm{p}} - E(\mathbf{X}_{\mathrm{n}} \mid \mathbf{R} = \mathrm{p}) \tag{A.10}$$

From A.9 and A.10 it is clear that the Poisson counting model predicts a negative linear relationship between RT and confidence whenever $v_{\mathrm{p}} + v_{\mathrm{n}}$ is constant (or relatively low in variance across observations). Note that $v_{\mathrm{p}} + v_{\mathrm{n}}$ is the rate parameter for the overall process. Inspection of A.6 and A.8 reveals that they can be rewritten in terms of θ_{p} and θ_{n}, where $\theta_{\mathrm{p}} = v_{\mathrm{p}}/(v_{\mathrm{p}} + v_{\mathrm{n}})$ and $\theta_{\mathrm{n}} = v_{\mathrm{n}}/(v_{\mathrm{p}} + v_{\mathrm{n}})$. Thus, response frequencies and confidence ratings depend on the relative rates at which positive and negative evidence accumulates. RT is also affected by the absolute overall rate. This is provides another way of understanding how RT might be lower for control ads even though $v_{\mathrm{p}T} > v_{\mathrm{p}C}$, as hypothesized. That is, $\theta_{\mathrm{p}T} > \theta_{\mathrm{p}C}$, but $v_{\mathrm{p}T} + v_{\mathrm{n}T} < v_{\mathrm{p}C} + v_{\mathrm{n}C}$ (i.e., evidence accumulation is slower for target ads than for control ads).

APPENDIX B: EXPERIMENT MATERIALS IN EXPERIMENT 4

	BELIEF VALENCE	
	POSITIVE	**NEGATIVE**
SUV	The Fiat Punto SUV has high quality.	The Fiat Punto SUV has low quality.
	The Fiat Punto SUV is a car I like.	The Fiat Punto SUV is a car not for me.
	The Fiat Punto SUV has a great price.	The Fiat Punto SUV has a price that's too high.
	The Fiat Punto SUV is cool.	The Fiat Punto SUV is boring.
	The Fiat Punto SUV is fast.	The Fiat Punto SUV is slow.
	The Fiat Punto SUV has good gas mileage.	The Fiat Punto SUV has bad gas mileage.
TV	The Panasonic Plasma has high quality.	The Panasonic Plasma has low quality.
	The Panasonic Plasma is a TV I like.	The Panasonic Plasma is a TV not for me.
	The Panasonic Plasma has a great price.	The Panasonic Plasma has a price that's too high.
	The Panasonic Plasma is cool.	The Panasonic Plasma is boring.
	The Panasonic Plasma is beautiful.	The Panasonic Plasma is ugly.
	The Panasonic Plasma has a clear picture.	The Panasonic Plasma has a fuzzy picture.
CHIPNUTS	The Fired-Up Chipnut has high quality.	The Fired-Up Chipnut has low quality.
	The Fired-Up Chipnut is a source I like.	The Fired-Up Chipnut is a source not for me.
	The Fired-Up Chipnut has a great price.	The Fired-Up Chipnut has a price that's too high.
	The Fired-Up Chipnut is cool.	The Fired-Up Chipnut is boring.
	The Fired-Up Chipnut is spicy.	The Fired-Up Chipnut is too spicy.
	The Fired-Up Chipnut has a delicious taste.	The Fired-Up Chipnut has an awful taste.
LOTION	The Nivea Body Shimmer Lotion has high quality.	The Nivea Body Shimmer Lotion has low quality.
	The Nivea Body Shimmer Lotion is a lotion I like.	The Nivea Body Shimmer Lotion is a lotion not for me.
	The Nivea Body Shimmer Lotion has a great price.	The Nivea Body Shimmer Lotion has a price that's too high.
	The Nivea Body Shimmer Lotion is cool.	The Nivea Body Shimmer Lotion is boring.
	The Nivea Body Silky Shimmer Lotion is effective.	The Nivea Body Silky Shimmer Lotion is ineffective.
	The Nivea Body Silky Shimmer Lotion has a pleasing fragrance.	The Nivea Body Silky Shimmer Lotion has a nasty smell.

REFERENCES

Alba, Joseph W., and Chattopadhyay, A. (1985). The effects of part-list cueing on attribute recall: Problem framing at the point of retrieval. In E. Hirschman & M. Holbrook (Eds.), *Advances in consumer research*, Vol. 12, Provo, UT: ACR, 410-413.

Alba, Joseph W. and Lynn Hasher (1983), "Is Memory Schematic?" *Psychological Bulletin*, 93 (March), 203-231.

Alba, Joseph W. and J. Wesley Hutchinson (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27(Sept), 123-56.

Bassili, John N. and Rick D. Brown (2005), "Implicit and Explicit Attitudes: Research, Challenges, and Theory." In *The Handbook of Attitudes*, eds. Dolores Albarracin, Blair T. Johnson, and Mark P. Zanna, Mahwah, NJ: Lawrence Erlbaum Associates, 543-74.

Berger, Ida E., and Andrew A. Mitchell (1989), "The Effect of Advertising on Attitude Accessibility, Attitude Confidence, and the Attitude-Behavior Relationship," *Journal of Consumer Research*, 16 (December), 269-79.

Bodur, H. Onur, David Brinberg, and Eloïse Coupey (2000), "Belief, Affect, and Attitude: Alternative Models of the Determinants of Attitude," *Journal of Consumer Psychology*, 9 (1), 17-28.

Brinol, Pablo, Richard E. Petty, and Zakary L. Tormala (2004), "Self-Validation of Cognitive Responses to Advertisements," *Journal of Consumer Research*, 30 (March), 559-73.

Burroughs, W. Jeffrey and Richard A. Feinberg (1987), "Using Response Latency to Access Spokesperson Effectiveness," *Journal of Consumer Research*, 14 (September), 295-99.

Cacioppo, John T. and Richard E. Petty (1981), "Social Psychological Procedures for Cognitive Response Assessment: The Thought Listing Technique," in *Cognitive Assessment*, cd. Thomas Merluzzi, Carol Glass, and Myles Genest, New York: Guilford, 309-42.

Cacioppo, John T., William von Hippel, and John M. Ernst (1997), "Mapping cognitive Structures and Processes Through Verbal Content: The Thought-Listing Technique," *Journal of Consulting and Clinical Psychology*, 65 (6), 928-40.

Chattopadhyay, Amitava and Joseph W. Alba (1988), "The Situation Importance of Recall and Inference in Consumer Decision Making," *Journal of Consumer Research*, 15 (Jan), 1-12.

Ericsson, K. Anders. and H. A. Simon (1980), "Verbal Reports as Data," *Psychological Review*, 82 (April), 215-51.

Fazio, Russell H. and Michael A. Olson (2003), "Implicit Measures in Social Cognition Research: Their Meaning and Use," *Annual Review of Psychology*, 54 (August), 297-327.

Fazio, Russell H., Martha C. Powell, and Carol J. Williams (1989), "The Role of Attitude Accessibility in the Attitude-to-Behavior Process," *Journal of Consumer Research*, 16 (December), 1-12.

Fazio, Russell H., David M. Sanbonmatsu, Martha C. Powell, and Frank R. Kardes (1986), "On the Automatic Activation of Attitudes," *Journal of Personality and Social Psychology*, 50 (February), 229-38.

Fishbein, Martin and Icek Ajzen (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, MA: Addison-Wesley.

Gardiner, John M. (1988), "Recognition Failures and Free-Recall Failures: Implications for the

Relation Between Recall and Recognition," *Memory & Cognition*, 16, 446-51.

Greenwald, Anthony G. (1968), "Cognitive Learning, Cognitive Response Persuasive, and Attitude Change," in *Psychological Foundations of Attitudes*, cd. Anthony G. Greenwald, Timothy C. Brock, and Thomas M. Ostrom, New York: Academic Press.

Greenwald, Anthony G, Debbie E. McGhee, and Jordan L. K. Schwartz (1998), "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test", *Journal of Personality and Social Psychology*, 74 (June), 1464-80.

Hawkins, Scott A. and Stephen J. Hoch (1992), "Low-Involvement Learning: Memory without Evaluation," *Journal of Consumer Research*, 19 (September), 212-25.

Johnston, William A., Veronica J. Dark, and Larry L. Jacoby (1985), "Perceptual Fluency and Recognition Judgments," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11 (January), 3-11.

Krishnan H. Shanker and Robert E. Smith (1998), "The Relative Endurance of Attitudes, Confidence, and Attitude-Behavior Consistency: The Role of Information Source and Delay," *Journal of Consumer Psychology*, 7 (3), 273-98.

Krugman, Herbert E. (1965), "The Impact of Television Advertising: Learning Without involvement," *Public Opinion Quarterly*, 29(Fall), 349-56.

Lutz, Richard J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," *Journal of Consumer Research*, 4 (March), 49-59.

MacInnis, Deborah J., Ambar G. Rao, and Allen M. Weiss (2002), "Assessing When Increased Media Weight of Real-World Advertisements Helps Sales," *Journal of Marketing Research*,

39 (November), 391-407.

Marks, Lawrence J. and Michael A. Kamins (1988), "The Use of Product Sampling and Advertising: Effects of Sequence of Exposure and Degree of Advertising Claim Exaggeration on Consumers' Belief Strength, Belief Confidence, and Attitudes," *Journal of Marketing Research*, 25 (Aug), 266-81.

McCulloch, Charles E. (2005), "Repeated Maesures ANOVA, R.I.P.?" Chance, 18 (3), 29-33.

Nisbett, Richard E. and Timothy DeCamp Wilson (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 84 (May), 231-59.

Novemsky, Nathan and Rebecca K. Ratner (2003), "The Time Course and Impact of Consumers' Erroneous Beliefs about Hedonic Contrast Effects," *Journal of Consumer Research*, 29 (March), 507-16.

Petty, Richard E., Pablo Brinol, and Zakary L. Tormala (2002), "Thought Confidence as a Determinant of Persuasive: The Self-Validation Hypothesis," *Journal of Personality and Social Psychology*, 82 (May), 722-41.

Petty, Richard E. and John T. Cacioppo (1979), "Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses," *Journal of Personality and Social Psychology*, 37 (December), 1915-26.

Petty, Richard E. and John T. Cacioppo (1986), *Communication and Persuasive: Central and Peripheral Routes to Attitude Change*, New York: Springer-Verlag.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Byes-Braem (1976), "Basic Objects in Natural Categories," *Cognitive Psychology*, 8 (July), 382-439.

Rosenthal, Robert and Ralph L. Rosnow (1991), *Essentials of Behavioral Research*, Boston, MA: McGraw-Hill.

Sanbonmatsu, David M., Frank R. Kardes, and Carol Sansone (1991), "Remembering Less and Inferring More: Effects of Time of Judgment on Inference about Unknown Attributes," *Journal of Personality and social Psychology*, 61 (October), 546-54.

Shavitt, Sharon and Timothy C. Brock (1986), "Self-Relevant Responses in Commercial Persuasive: Field and Experimental Test," In *Advertising and Consumer Psychology*, ed. Jerry Olson and Keith Sentis. Praeger Publishers: New York, 149-71.

Shrum, L. J., Robert S. Wyer., and Thomas C. O'Guinn (1998), "The Effects of Television Consumption on Social Perceptions: The Use of Priming Procedures Investigate Psychological Processes," *Journal of Consumer Research*, 24 (March), 447-58.

Smith, Edward. E., Edward J. Shoben, and L. J. Rips (1974), "Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions," *Psychological Review*, 81 (3), 214-41.

Smith, Philip. L. and Douglas Vickers (1988), "The Accumulator Model of Two-Choice Discrimination," *Journal of Mathematical Psychology*, 32 (June), 135-68.

Townsend, James T. and F. Gregory Ashby. (1983), *The Stochastic Modeling of Elementary Psychological Processes*, New York: Cambridge University Press.

Toy, Daniel R. (1982), "Monitoring Communication Effects: A Cognitive Structure/Cognitive Response Approach," *Journal of Consumer Research*, 9 (June), 66-76.

Van Zandt, Trisha (2000), "ROC Curves and Confidence Judgments in Recognition Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (May), 582-600.

Van Zandt, Trisha and Mildred M. Maldonado-Molina (2004), "Response Reversal in Recognition Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30 (November), 1147-66.

Wansink, Brian, Michael L. Ray, and Rajeev Batra (1994), "Increasing Cognitive Response Sensitivity," Journal of Advertising, 23(June), 65-75.

Whittlesea, Bruce W. A., and Richard L. Wright (1997), "Implicit (and explicit) learning: acting adaptively without knowing the consequences," *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21 (January), 181-200.

Williams, Patti, Gavan J. Fitzsimons, and Lauren G. Block (2004), "When Consumers Do Not Recognize 'Benign' Intention Questions as Persuasive Attempts," *Journal of Consumer Research,* 31 (December), 540-50.

Wolfinger, R. D., and M. Chang, (1998). Comparing the SAS GLM and MIXED procedures for repeated measures. Cary, NC: SAS Institute Inc.

Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," *Journal of Marketing Research*, X (February), 53-62.

Wyer, Robert S. and Dolores Albaraccin (2005), "Belief Formation, Organization, and change: Cognitive and Motiviational Influences," in *The Handbook of Attitudes*, eds. Dolores Albarracin, Blair T. Johnson, and Mark P. Zanna, Mahwah, NJ: Lawrence Erlbaum Associates, 273-322.

Exhibit 1
EXPERIMENTAL HYPOTHESES

H1: More positive responses will be observed in thought recognition and belief verification tasks for high thought likelihood (target) advertisements than for low thought likelihood (control) advertisements.

H2: Reaction times for positive responses to thought recognition and belief verification tasks will be lower for high thought likelihood (target) advertisements than for low thought likelihood (control) advertisements.

H3: Confidence ratings for positive responses to thought recognition and belief verification tasks will be higher for high thought likelihood (target) advertisements than for low thought likelihood (control) advertisements.

H4: When the parameters of the counting model of responses, reactions times, and confidence ratings are estimated, v_{pT} will be greater than v_{pC} .

H5: When the parameters of the counting model of responses, reactions times, and confidence ratings are estimated, v_{pT} - v_{pC} will be greater than v_{nT} - v_{nC}.

H6: The valence of a specific cognitive response has a positive effect on attitude, assuming that the thought recognition or belief verification response was positive.

H7: Reaction times for positive responses to thought recognition and belief verification tasks have (a) a positive effect on attitude (b) that is moderated by valence such that the effect is positive only when valence is positive.

H8: Confidence ratings for positive responses to thought recognition and belief verification tasks have (a) a positive effect on attitude (b) that is moderated by valence such that the effect is positive only when valence is positive.

H9: The effects of valence, reaction time, and confidence on attitude (i.e., H6, H7, and H8) are reversed when the associated thought recognition or belief verification response is negative.

Table 1
SUMMARY OF MEASURES RELATED TO COGNITIVE RESPONSE

Task	Sample Stimulus	Measure	Type	References
Thought Listing	We are now interested in everything that went through your mind when viewing the magazine ads earlier in this experiment. Please list these thoughts, whether they were about yourself, the ads, or the products; whether they were positive, neutral and/or negative.	Response (open-ended)	Explicit	Cacioppo & Petty (1981); Cacioppo et al. (1997); Ericcson & Simon (1980); Toy 1982; Wansink, Ray, & Batra (1994)
		Valence (& other coding schemes)	Explicit or Implicit	Chattopadhyay & Alba (1988); Cacioppo & Petty (1981)
		Response Time	Implicit	None found.
		Confidence	Explicit	Brinol, Petty, & Tormala (2004); Petty, Brinol, & Tormala (2002)
Thought Recognition	When viewing advertisements, people often have a variety of thoughts and reactions. Regardless of what you believe now, when viewing the magazine ads earlier in this experiment, do you ever think to yourself, 'The Fiat SUV is fast'? *	Response (Yes/No)	Explicit	Cacioppo et al. (1997)
		Valence (& other coding schemes)	Explicit or Implicit	Cacioppo et al. (1997)
		Response Time	Implicit	Van Zandt & Maldonado-Molina (2004)
		Confidence	Explicit	Krishnan & Smith (1998); Van Zandt (2000)
Belief Verification	The Fiat SUV is fast. *	Response (Agree/Disagree)	Implicit	Fishbein & Ajzen 1975; Lutz 1975
		Valence (& other coding schemes)	Explicit or Implicit	Bodur, Brinberg, & Coupey (2000); Marks & Kamins (1988)
		Response Time	Implicit	Smith, Shoben, & Rips (1974); Rosch et al. (1976)
		Confidence	Implicit	Bodur, Brinberg, & Coupey (2000); Krishnan & Smith (1998); Marks & Kamins (1988)

* The underlined word was shown after a delay. Response times were measured from the onset of that terminal word.

Table 2
LEAST-SQUARE MEAN RESPONSE TIMES AND CONFIDENCE RATINGS FOR EXPERIMENTS 1 - 3

Thought Likelihood	Response Type	Frequency Chosen	Percent Chosen	Response Time (msec)	Confidence Rating
		Experiment 1 - Thought Recognition			
Low	Positive (Yes)	135	45	2789	1.80
High	Positive (Yes)	211	73	2391	2.12
Difference			-28	398	-.32
Low	Negative (No)	164	55	2871	1.95
High	Negative (No)	80	27	2971	1.90
Difference			28	-100	.05
		Experiment 2 - Belief Verification			
Low	Positive (Agree)	223	75	2450	1.60
High	Positive (Agree)	263	99	1955	2.05
Difference			-13	495	-.45
Low	Negative (Disagree)	73	25	2701	1.00
High	Negative (Disagree)	35	12	2269	1.21
Difference			13	432	-.21
		Experiment 3 - Thought Recognition			
Low	Positive (Yes)	75	48	2764	1.78
High	Positive (Yes)	115	70	2404	2.10
Difference			-22	360	-.32
Low	Negative (No)	82	52	2690	2.10
High	Negative (No)	50	30	3360	1.80
Difference			22	-670	.30
		Experiment 3 - Belief Verification			
Low	Positive (Agree)	88	45	2968	1.45
High	Positive (Agree)	102	65	2270	2.00
Difference			-20	698	-.55
Low	Negative (Disagree)	25	55	4095	-.61
High	Negative (Disagree)	16	35	2958	1.01
Difference			20	1137	1.65

Table 3
ESTIMATED PARAMETERS FOR A POISSON COUNTING MODEL OF RESPONSE PROBABILITIES, REACTION TIMES, AND CONFIDENCE RATINGS

Stimuli	N	K_p	K_n	T_0	$v_{p\mathrm{T}}$	$v_{n\mathrm{T}}$	$v_{p\mathrm{T}} + v_{n\mathrm{T}}$	$v_{p\mathrm{C}}$	$v_{n\mathrm{C}}$	$v_{p\mathrm{T}} + v_{n\mathrm{T}}$	*VAF**	D_1	t_1	D_2	t_2
Experiment 1 - Thought Recognition															
Lotion	147	4	4	-.2	.9	.8	1.7	.6	1.2	1.8	.06 / .37	.3	5.4	-.3	-5.2
Chipnuts	149	4	6	-.2	1.5	.9	2.4	1.3	1.3	2.6	.14 / .66	.2	2.0	-.4	-5.1
SUV	147	5	6	-.1	1.3	1.2	2.5	.7	1.9	2.5	.14 / .63	.6	8.4	-.6	-7.3
TV	147	5	6	-.3	2.0	.5	2.5	1.6	.8	2.5	.21 / .65	.3	3.4	-.3	-4.8
Experiment 2 - Belief Verification															
Lotion	149	4	4	.5	1.4	.7	2.1	1.3	1.0	2.3	.16 / .77	.2	2.0	-.3	-4.7
Chipnuts	148	4	4	-.4	2.1	.6	2.8	1.9	.9	2.8	.03 / .32	.3	2.1	-.3	-3.5
SUV	148	5	5	-.6	1.7	.9	2.6	1.1	1.1	2.2	.16 / .76	.6	6.9	-.2	-3.6
TV	149	5	5	-.1	3.1	.7	3.8	2.4	.7	3.1	.25 / .97	.7	4.5	-.0	-.4
Experiment 3 - Thought Recognition															
Lotion	80	6	6	-.6	1.0	1.1	2.0	.7	1.4	2.1	.20 / .65	.3	3.5	-.3	-3.3
Chipnuts	81	5	6	-.5	1.8	.8	2.6	1.3	1.3	2.6	.21 / .80	.5	4.0	-.5	-4.7
SUV	81	5	6	-.2	1.4	1.3	2.7	.7	1.9	2.6	.05 / .22	.7	7.0	-.6	-4.8
TV	80	5	6	-.6	1.5	.4	1.9	1.4	.7	2.1	.16 / .58	.1	.6	-.3	-4.6
Experiment 3 - Belief Verification															
Lotion	56	5	6	.6	2.0	1.1	3.1	1.2	1.1	2.2	.34 / .90	.8	4.7	.1	.4
Chipnuts	60	6	6	.0	3.3	.8	4.1	2.5	1.1	3.6	.25 / .73	.8	3.5	-.3	-2.2
SUV	56	5	6	-.0	1.4	1.2	2.6	1.1	1.2	2.4	.15 / .74	.3	2.2	-.0	-.3
TV	59	5	6	-.1	2.5	.4	2.9	2.2	.8	3.0	.28 / .78	.3	1.3	-.4	-3.5

K(.) - Decision threshold; T_0 - Response time intercept in seconds; *V*(.) - Rate parameters

p - Positive Response ("Yes" or "Agree"); *n* - Negative Response ("No" or "Disagree"); T - Target Ad ; C - Control Ad

*VAF** - Variance Accounted For: Overall/Systematic (see Appendix A for definitions)

D_1 - $v_{p\mathrm{T}}$ - $v_{p\mathrm{C}}$, predicted to be greater than 0 (H4)

R_2 - $v_{n\mathrm{T}}$ - $v_{n\mathrm{C}}$, predicted to be less than D_1 (H5)

Table 4
REGRESSION COEFFICIENTS FOR ATTITUDES IN THE BELIEF VERIFICATION TASK IN EXPERIMENT 3

Response	Valence	Reaction Time (Positive Valence)	Reaction Time (Negative Valence)	Confidence (Positive Valence)	Confidence (Negative Valence)
Positive	.21[a]	-.08[b]	.04[c]	.22[d]	.50[e]
Negative	.13[f]	.15[f]	.54[f]	-.59[f]	-.68[f]

[a] H6 predicts this coefficient to be positive.

[b] H7a predicts this coefficient to be negative.

[c] H7b predicts this coefficient to be positive.

[d] H8a predicts this coefficient to be positive.

[e] H8b predicts this coefficient to be negative.

[f] H9 predicts these coefficients to be reversed in sign compared to the coefficient above them.

Table 5
REGRESSION MODEL RESULTS FOR EXPERIMENT 4

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	β	*F*	*β*	*F*	*β*	*F*	*β*	*F*	*β*	*F*	*β*	*F*
Ad Dummy Variables												
Lotion - Target	---	---					---	---				
Chipnuts - Target	-1.25	46.6					-.82	21.1				
SUV - Target	.20	1.2					.38	4.9				
TV - Target	.85	19.9					.29	2.7				
Lotion - Control	.32	2.5					.07	.2				
Chipnuts - Control	-1.25	47.9					-.83	21.8				
SUV - Control	-.65	12.0					-.17	.9				
TV - Control	1.00	29.4					.22	1.5				
Product Dummy Variables												
Lotion			---	---					---	---	---	---
Chipnuts			-1.41	124.8					-.86	44.9	-.71	30.1
SUV			-.35	7.5					.08	.4	.09	.5
TV			.76	34.2					.22	3.0	.30	5.5
Belief Verification Indices												
Valence					.35	330.0	.27	173.4	.27	170.0	.31	241.0
RT (Agree)					-.12	4.5	-.10	3.2	-.11	4.3		
RT (Disagree)					.02	.2	.04	.7	.04	.6		
Confidence (Agree)					.16	5.6	.30	21.9	.31	22.1		
Confidence (Disagree)					-.14	3.5	-.04	.4	-.05	.6		
Fit Measures												
-2LL	2218		2246		1706		1610		1622		1667	
AIC	2238		2258		1722		1640		1644		1681	
BIC	2274		2279		1749		1691		1682		1705	

* Dummy variables are 0/1. The Valence Index is the number of positive agree responses minus the number of negative agree responses. The RT index is the average positive RT minus the average negative RT (in seconds). The Confidence Index is the average positive confidence rating minus the average negative confidence rating (ratings were a 3-point scale). See text for computational details.

FIGURE 1

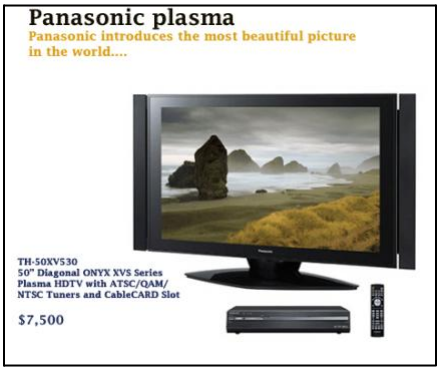
ADVERTISEMENTS USED IN ALL EXPERIMENTS

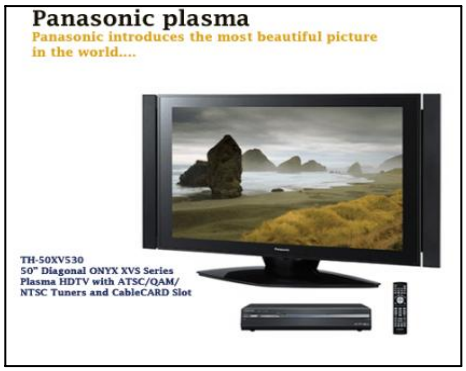
Taraet Attributes ***Taraet Ads***

Fast

Spicy

Expensive





Cheap

FIGURE 2

An Example of the Poisson Counting Model of Decisions, Response Times, and Confidence

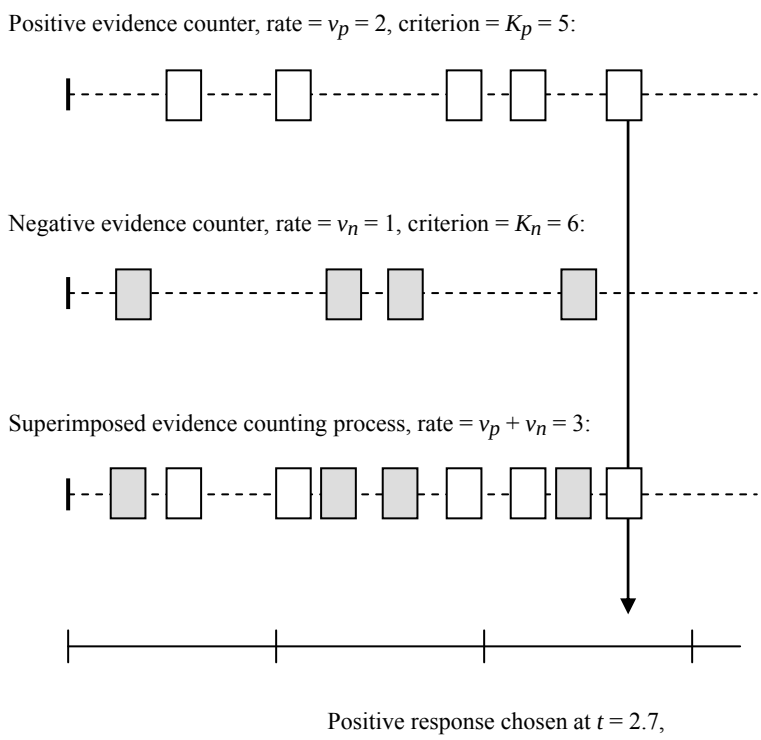
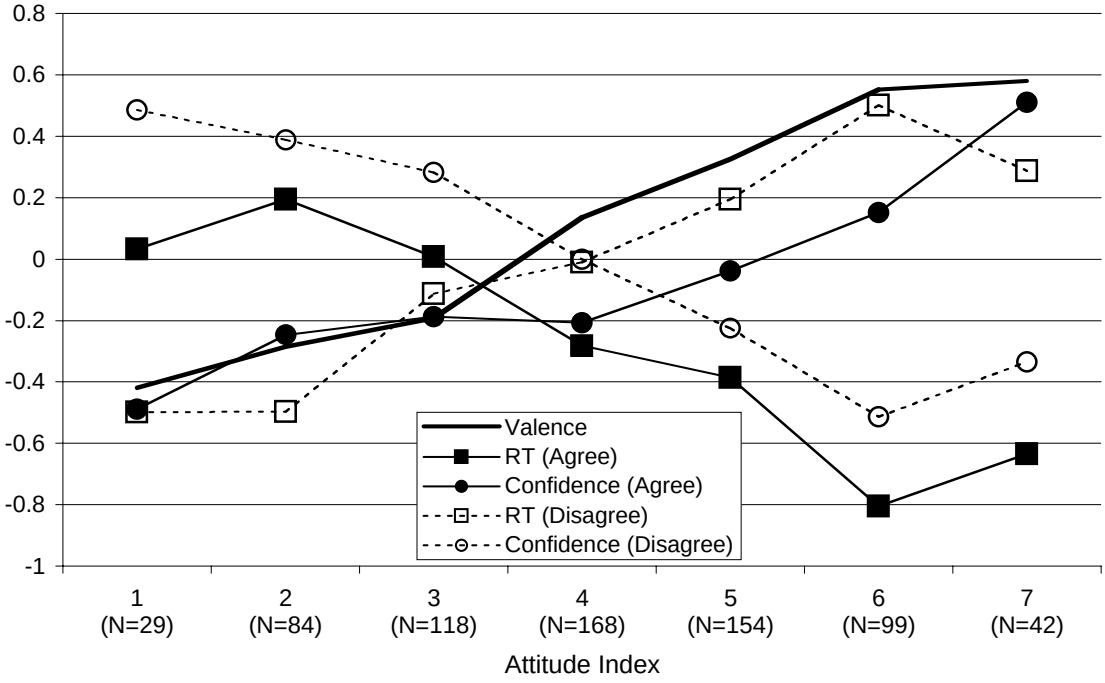


FIGURE 3

Belief Verification Indices Plotted against Attitude



View publication stats