A Meta-Analysis of the ELM’s Argument Quality × Processing Type Predictions

Christopher J. Carpenter
Department of Communication, Western Illinois University, Macomb, IL 61455, USA

One of the key dual-process model predictions is that audiences will be more persuaded by strong persuasive arguments than weak and that this difference in persuasiveness will be larger when they are processing centrally rather than peripherally. A series of meta-analyses were conducted (k = 134) to assess this claim and explore moderators. The data were generally consistent with the hypothesized interaction. The effects tended to be smaller when pre-post designs were used rather than post-test only. Assessments of the strength of the inductions did not tend to be associated with the size of the effects associated with those inductions.

Keywords: Meta-Analysis, Persuasion, Dual-Process, Argument Quality, Elaboration Likelihood Model, Heuristic-Systematic Model.

doi:10.1111/hcre.12054

The elaboration likelihood model (ELM) of persuasion (Petty & Cacioppo, 1981, 1986) and the somewhat similar dual-process heuristic-systematic model (HSM; Chaiken, 1980) have come to dominate persuasion research. By distilling the myriad variables and processes that were being studied into a single continuum with two kinds of processing, the field of persuasion research was able to find coherence that had been previously lacking. The heuristic value of dual-process models have been enormous with hundreds of studies testing various aspects of these theories, using them to explain previous findings, and exploring new persuasion variables.

Yet, dual-process models and the ELM in particular have been criticized on many fronts. The ELM has been called nonfalsifiable (Stiff & Boster, 1987), accused of creating needless neologisms (Hamilton, Hunter, & Boster, 1993), described as useless for designing persuasive messages (Bittner & Obermiller, 1985), and reproached for being too vague to study systematically (Mongeau & Stiff, 1993). There is even some uncertainty about which labs can produce the predicted effects and which cannot (Johnson & Eagly, 1989). In addition, some researchers have reported studies that ostensibly falsify the theory (Chung, Fink, & McLaughlin, 2011; Mitchell, 2000; Mongeau & Williams, 1996; Park, Levine, Westerman, Orfgen, & Foregger, 2007; van Dijk, 2004).

---

Corresponding author: Christopher J. Carpenter; e-mail: Cj-carpenter2@wiu.edu
Hustinx, & Hoeken, 2003; Ziegler, Diehl, & Ruther, 2002). On the other hand, there are also many successful replications. The inconsistency requires explanation and the extent of it, quantification.

When the literature concerning a theory is uncertain, the best solution is to conduct a meta-analysis. As Hunter and Schmidt (2004) explained, a meta-analysis produces weighted averages of the effects predicted by the theory. Sampling error will naturally produce many failed replications due to the small samples common in the social sciences. Meta-analysis can help remove the bias produced by statistical significance testing and provide an estimate of the effects associated with the theory’s predictions. Perhaps more importantly, a meta-analysis can help determine what causes the effects predicted by the theory to appear in some cases and not others by looking at between study moderators of the effects. In the case of the ELM, many questions have been raised about the impact of various methodological choices (Hamilton et al., 1993; Johnson & Eagly, 1989; Petty, Kasmer, Haugtvedt, & Cacioppo, 1987). To help shed some light on these questions and suggest additional avenues for fruitful research, this article reports the results of a series of meta-analyses focused on one of the predictions of the ELM and the HSM. Dual process theory will be briefly reviewed in order to describe the predictions being tested in the meta-analyses. Then the various moderators tested in these meta-analyses will be discussed.

A brief overview of the elaboration likelihood model

Although the predictions tested in these meta-analyses represent an area of overlap between the HSM and the ELM, the ELM will be focused on here due to the higher proportion of articles and the greater controversy associated with it. The ELM was described most fully in Petty and Cacioppo’s seminal 1986 monograph. At the core of the model is the elaboration continuum that ranges from no cognitive elaboration of the issues relevant to the attitude object in the persuasive message to completely elaborating all of the issue-relevant information in the message. Someone can completely ignore a persuasive message, they can think about it superficially, or they can carefully examine all of its content. Petty and Cacioppo described larger amounts of cognitive elaboration as including processes such as attempting to comprehend the information clearly, weighing information, comparing the information in the message to previously known information, and considering the likelihood of the claims made.

Petty and Cacioppo (1986) described the low end of the continuum as peripheral processing. Peripheral processing includes any means of arriving at an attitude judgment without careful considering of the issue-relevant information. One peripheral process is the application of simple judgment heuristics (Chaiken, 1980) such as “if the source is credible, I should agree with her or him” or “if the speaker has many arguments, her or his claims are likely to be true.” According to the ELM authors, other peripheral processes can include those described by balance theory (Heider, 1958) and self-perception theory (Bem, 1972).

The high processing end of the continuum is called central processing (Petty & Cacioppo, 1986). These processes include any cognitive process that involves careful
examination of the evidence presented in a persuasive message. They argued that the processes described by inoculation theory (Lumsdaine & Janis, 1953) and the theory of reasoned action (Fishbein & Ajzen, 1975) might occur under central processing.

The key hallmark of central processing (and the focus of this investigation) is that a centrally processing audience will likely be more persuaded by strong arguments than weak arguments. By carefully considering the message information, the centrally processing audience is able to detect when the evidence provided by the message is weaker and when it is stronger.

An audience that is peripherally processing is predicted to be substantially less likely to detect differences in argument quality. Because a peripherally processing audience is only superficially examining the message content, they are unlikely to be devoting enough cognitive resources to the message to notice when it is weak or strong.

As a result, the theory predicts a nonadditive effect of argument quality on persuasion such that strong arguments will produce more positive attitudes than weak under central processing, but very little difference under peripheral processing. Specifically, in central processing, there will be a substantial argument strength effect, and in peripheral processing, there will be either a null or a small effect of argument strength. The comparison of these two effects represents the key test of these predictions. The theory also predicts when central or peripheral processing will occur.

The ELM argues that to process centrally, someone must be able and motivated to process the issue-relevant information (Petty & Cacioppo, 1986). Ability to process can be influenced by a number of factors. For example, one must not be distracted (Petty, Wells, & Brock, 1976) and the message must not be too confusing (Hafer, Reynolds, & Obertynski, 1996). Theoretically, any factor that causes someone to be unable to attend to and think carefully about the message will reduce elaboration likelihood.

To process centrally, one must also be motivated to do so (Petty & Cacioppo, 1986). The most commonly studied means of affecting motivation to process are outcome-relevant involvement inductions (e.g., Petty, Cacioppo, & Goldman, 1981). When topics of the persuasive messages directly affect the ability of the audience to reach their goals, they are considered high outcome-relevant involvement issues and as such spur central processing (Johnson & Eagly, 1989). When the topics do not affect the audience, they are considered low involvement issues and peripheral processing is thought to be the result.

For example, Petty et al. induced high outcome-relevant involvement in a college student audience by informing them that the university was considering instituting a policy that would require the students to pass a comprehensive exam in order to graduate. The low-involvement audience were told that the university was considering implementing the policy in 10 years and thus not affecting the student participants’ chances of graduating. In addition to motivation based on the issue, Cacioppo and Petty (1982) proposed that some people vary in the extent to which they enjoy
complex message processing such that some people are high in need for cognition
and are likely to centrally process most of the persuasive messages to which they are
exposed.

The theory also makes several other predictions about attitude strength and biased
processing (Petty & Cacioppo, 1986). It has more recently been expanded to include
meta-cognitions (Petty, Briñol, & Tormala, 2002). These additional aspects of the the-
ory are not relevant to the current investigation. The original theory does have an
additional postulate that raises serious concerns about the possibility of even con-
ducting a meta-analysis of the ELM and has generated much of the controversy about
the theory.

Specifically, in their first full description of the theory, Petty and Cacioppo (1986)
included the “multiple roles postulate.” Some authors have called for the theory to
attempt to a priori classify persuasive variables into peripheral cues (cues that impact
persuasion under peripheral processing), central cues (cues that impact persuasion
under central processing), or cues that affect the extent of elaboration (Stiff & Boster,
1987). Instead, the theory states that any given cue can take on any of the specified
roles. Petty and Cacioppo’s (1986) example is the attractiveness of the model in a
magazine ad. In an ad for an automobile, the attractiveness of the source would be
a peripheral cue as it is not related to the issue. On the other hand, in a shampoo ad,
if the model has particularly nice looking hair, her hair would be relevant informa-
tion and could affect the persuasiveness of the ad when the audience is processing
centrally. Critics such as Stiff and Boster (1987) argued that if any cue can take on
any role in the ELM, then any result would be consistent with the theory and thus the
theory could not be falsified.

This issue raises a problem for conducting meta-analyses that test the theory’s pre-
dictions. A meta-analysis of the effects of a given cue like source credibility would face
the difficulty of trying to decide if source credibility should only have influenced per-
suasion when the audience is processing centrally, only when processing peripherally,
or both. The ELM provides the guidance that such a cue would be a central cue when
it is relevant to the issue at hand, but determining what is relevant and what is not is
difficult. Furthermore, Petty and Cacioppo (1986) noted that relevance is likely to be
a subjective judgment and vary between people. Also, why Petty et al. (1987) would
claim that the attractiveness of the source is a central cue for considering shampoo,
whereas possessing expertise in education topics would not be a central cue when the
issue concerns academic policy remains unclear (e.g., Petty et al., 1981).

Multiple roles hypothesis
There is one prediction made by the ELM that does not seem affected by the multiple
roles hypothesis: the impact of argument quality. Petty and Cacioppo (1986) make it
clear that argument quality is only a central cue. They explain that other message fac-
tors such as message length and number of arguments may be central or peripheral
cues. The manipulated quality of the arguments in the message, however, is only a cen-
tral cue. This prediction is the focus of the meta-analyses reported here. Specifically,
the size of the effect that argument quality has on postmessage attitudes should be substantially smaller when the audience is processing peripherally than when they are processing centrally.

The other corollary of that prediction is that when the arguments are strong, a centrally processing audience is predicted to have more positive (message-congruent) attitudes than a peripherally processing audience. This effect would result because the centrally processing audience would notice and appreciate the higher-quality arguments and thus be more persuaded, whereas the peripherally processing audience would not. When the arguments are weak, the centrally processing audience is predicted to have more negative (message-incongruent) attitudes than a peripherally processing audience. This effect suggests that peripherally processing audiences are not as able to notice how weakly the poor arguments support the argument, whereas a centrally processing audience would.

Although the HSM and the ELM agree on the predicted Argument Strength × Processing Type interaction, there are several other areas in which the two theories make divergent predictions. Chaiken and Stangor (1987) note that the HSM distinguishes additional types of involvement, whereas the ELM predicts that all message involvement produces the same effects. In addition, Chaiken and Stangor explain that although the ELM predicts that peripheral cues have no effect on persuasion under central processing, the HSM allows for parallel processing in which cues such as source credibility may also be used by the audience to form judgments. These differences do not, however, affect the interaction analyzed in this investigation.

**Proposed moderators**

**Sources**

One of the most surprising findings from Johnson and Eagly’s (1989) meta-analysis of the interaction between outcome-relevant involvement and argument quality was the finding that their predicted interaction only tended to emerge when Ohio State University (OSU) researchers conducted the study. They found that methodological moderators could not explain the finding. Instead, they suggested that a form of publication bias was operating such that OSU researchers were less likely to publish failed replications than non-OSU. Petty and Cacioppo (1990) argued that perhaps some of the failed replications were attributable to methodological differences. Given that there have been many new studies published since their debate, meta-analysis can assess some of these claims.

First, the presence of “Ohio State effect” can be re-examined to determine if OSU researchers find larger effects in the predicted direction than non-OSU. In addition, the trim-and-fill algorithm (Duval & Tweedie, 2000) can be used to assess publication bias by estimating the extent to which a body of studies is missing studies with null findings. This algorithm examines the distribution of the effects, notes deviation from the distribution expected by sampling error, and estimates what the effect size would be if the theoretically missing studies were included. If the trim-and-fill algorithm finds greater publication bias among the body of OSU
studies than non-OSU, such a finding would be consistent with Johnson and Eagly’s speculation that OSU researchers did not submit failed replications for publication. In addition, meta-analysis can determine if various methodological moderators such as using the methods preferred by Petty and Cacioppo (1986) to pretest argument quality or other methodological variations impact the average effect sizes.

In addition, there are three academic disciplines that have produced the bulk of the ELM research: social psychology, marketing/advertising, and communication. These three disciplines tend to take somewhat different approaches to researching persuasion and it may be that the effects obtained in ELM research varies among them. Therefore, this meta-analysis will examine the field from which the research came as an additional moderator.

Study designs
Other ELM critics argued that the theory only explains attitude formation rather than change (Hamilton et al., 1993). Essentially, the ELM predictions only work when an audience is encountering a new object and forming an attitude for the first time. For example, in the Petty et al. (1981) study, most of the students had probably never heard of a university requiring a comprehensive exam for graduation and, thus, they formed a new attitude for the first time when hearing the persuasive message. Attitude change occurs when the audience is persuaded by a message about an issue for which they already have an attitude.

To establish that actual attitude change has occurred, attitudes must be measured before exposure to the persuasive message and again afterward. Thus, the actual change in attitude scores can be quantified with such a pre- and posttest design. Most persuasion studies, however, use a post-test only such that attitudes are only measured after exposure to the persuasive message and any variation in pretest attitudes is thought to be taken into account by random assignment to conditions.

If there is a boundary condition for the ELM such that it does not apply to attitude change, then studies testing the theory that used a pre- and posttest design would find substantially smaller effects than posttest-only designs. The majority of studies in this literature use posttest-only designs so evidence that studies that measure actual change in attitudes are an important check to determine if attitude change is occurring. Comparing the two designs across studies meta-analytically can examine this claim.

Induction types
A variety of other moderators are potentially of interest to persuasion researchers as well. One of the core propositions of the ELM is that reducing the ability of the audience to cognitively process the message carefully and reducing their motivation to do so produce the same kind of effects (Petty & Cacioppo, 1986). Such an insight is beneficial to the field of persuasion research as it reduces the number of variables and thus advances research. The ELM predicted that the effect produced by ability inductions should be comparable to the effect of motivation inductions. Such a prediction can be
tested meta-analytically by comparing the effects associated with the different kinds of inductions.

Given the large number of studies available, determining if there are differences in the effects produced by specific inductions of peripheral and central processing is also possible. These included the already mentioned outcome-relevant involvement and need for cognition. They also included studies that argued that people process centrally when in negative moods and peripherally when in positive moods (e.g., Worth & Mackie, 1987). There were also studies that measured participants’ personalities and matched messages they thought would cause people with those personality types to process centrally and peripherally (e.g., DeBono & Harnish, 1988). The effects produced by these four processing type inductions will be compared to each other.

*Induction checks and pilot tests*

Many of the studies presented also offered information about induction checks and pilot tests of some of the key variables. These checks and tests will be examined as potential moderators. The method that the various studies used for pilot testing the argument quality inductions might produce variations in the associated effect sizes if those methods are producing qualitatively different types of argument quality inductions. Some studies used simple semantic differential and likert scales and if the difference between the strong and the weak arguments is statistically significant, they used the messages. Other studies pilot tested by having the audience list their thoughts about the messages and they used arguments that produce a sufficiently high proportion of positive to negative thoughts for strong and vice versa for weak arguments. Still others focused on inducing specific aspects of likelihood or logic of the message.

A large group of studies simply used the same argument quality induction that was provided in detail by Petty and Cacioppo (1986) that offered strong and weak arguments in favor of comprehensive exams. These different kinds of argument quality inductions would theoretically be associated with variation in the effect size associated with argument quality when the audience is processing centrally, but not peripherally. Peripherally processing audiences are predicted to be indifferent to variation in argument quality inductions, whereas the centrally processing audiences are the ones predicted to be responsive to argument quality.

In addition, the strength of the argument quality induction would be theoretically likely to be associated with a greater increase in persuasiveness for strong over weak arguments when the audience is processing centrally but not peripherally. Theoretically, a stronger argument quality induction would be likely to produce stronger argument quality effects for central processors. To determine how strong the argument quality inductions in the available studies were, the effect sizes for either the induction checks or the pilot tests will be calculated. The effect sizes associated with both argument quality pilot tests and induction checks will be examined as potential predictors of argument quality effects in the primary effect sizes.

Some of the studies also provided information about the size of the processing type induction via induction checks. Again, the ELM predicts that a larger processing type
induction would be likely to cause a larger difference in the effect size of argument quality for central over peripheral processors. That hypothesis will also be tested.

Summary
This report will describe a series of meta-analyses of the ELM prediction that strong arguments will be substantially more persuasive than weak arguments when the audience is processing centrally and not when they are processing peripherally. Five effect sizes will be calculated. First, the difference between the persuasiveness of strong versus weak arguments when the audience is processing peripherally is predicted by the theory to be nil or very small. The difference between strong and weak arguments when they are processing centrally is predicted to be ample. The effect size for argument strength under peripheral processing will be subtracted from the effect size under central processing. The difference between those two effect sizes show the increase predicted by the ELM such that a larger value is more theory consistent. The final two effect sizes were calculated to determine if it is only strong arguments to which the centrally processing audiences are more likely to react, weak arguments only, or, as the theory predicts, both. If the average effect for these two is about equal, that data would be consistent with the theory. The third of these, the difference in the effect size for argument quality between centrally processing and peripherally processing audiences represents the core of the ELM position. The ELM predicts that the effect of argument quality on postmessage attitudes will be stronger in central processing conditions than peripheral (Petty & Cacioppo, 1986).

Method
Sample of studies
The search terms “elaboration likelihood model” and “heuristic-systematic model” were successively used to search for studies in Google Scholar. In addition, the same engine was used to search for articles citing the seminal ELM (Petty & Cacioppo, 1986) and HSM (Chaiken, 1980) studies. In this case, the HSM makes the same predictions as the ELM. Rather than search for aspects of the studies desired, searching for the two primary persuasion dual-process theories ensured that studies were designed to test the targeted types of processing. Although previous meta-analyses such as Stiff (1986) and Johnson and Eagly (1989) targeted only involvement, this meta-analysis targeted central and peripheral processing generally as Petty and Cacioppo conceived it. Studies that were not intentionally inducing these types of processing risk introducing unnecessary variance. A message was also sent to the Communication, Research, and Theory Network (CRTNET) e-mail listserv requesting unpublished studies.

Several criteria were used for inclusion/exclusion of studies. First, the study had to be attempting to induce or measure predictors of central and peripheral processing. Some studies proposed unusual means of inducing the two types of processing that were not included in the original descriptions of the theory. In order for nonstandard means of inducing processing to be included, they had to demonstrate successful
induction checks such as the differential impact of thought valence on attitudes (Petty & Cacioppo, 1986). In addition, the study had to test a factorial experimental design in which those types of processing were induced and crossed with an argument quality induction with strong and weak arguments. Finally, the article had to report enough information to derive the desired effect sizes. The authors of articles that were otherwise fit for inclusion but did not provide information were e-mailed to request the information. The result was that there were 134 studies included with a combined sample of $N = 15,426$. The included articles are listed in an Appendix that accompanies the electronic version of this article.

Coding
In addition to the effect sizes, several aspects of each study were coded. Aspects of the articles included the year it was published, in which of the three fields it was published (social psychology, marketing/advertising, and communication), and if the authors were affiliated with OSU. For this last moderator, Johnson and Eagly’s (1989) criteria were used such that studies authored by Petty, Cacioppo, or an OSU student of Petty’s were coded as an OSU study.

Several study characteristics were coded. The use of an ability or motivation induction of central versus peripheral processing was recorded as well as the specific type of processing induction. In some cases, both types were used simultaneously and in one case, central versus peripheral processing was assessed by measuring thought-listing. The means by which the strong versus weak arguments were developed was also recorded such that the authors could have used a thought-listing pilot test, a pilot test employing continuous rating scales, the induction of message variables, or the use of Petty and Cacioppo’s (1986) comprehensive exam arguments. Whether the study used a pre- and posttest design or only a postmessage design to measure attitudes was also noted.

The effect sizes for the results of pilot tests and induction checks were also calculated. If a pilot test with a separate sample to establish argument quality was described, the effect size associated with the difference in perceived argument quality in that sample was calculated and recorded. The effect sizes for any provided induction checks for processing type and argument quality were also calculated and recorded. The reliability of the attitude measure was also recorded. There were 86 internal consistency reliabilities reported, ranging from .7 to .98. There were 48 studies that did not report reliability information for the attitude measures.

Meta-analysis method
The method of estimating the weighted average effects followed Hunter and Schmidt’s (2004) recommendations. Their method has emerged as both the most commonly used method in communication research and the most valid (Anker, Reinhart, & Feeley, 2010). A correlation coefficient was calculated for each effect size. In order to reduce the effects of sampling error on the average effect size estimate, the studies were weighted by their sample size. In order to reduce the effects of measurement
error on the estimate, Hunter and Schmidt’s artifact distribution method was used to correct the estimates.

Next, the heterogeneity of variance was assessed. According to Hunter and Schmidt (2004), a certain amount of variation in the effect in the estimate of a relationship across studies is to be expected from sampling error. They provided a formula for estimating the expected variance that could be attributed to sampling error. They argued that if one calculates a ratio of the variance predicted from sampling error and measured artifacts in the numerator and the variance observed among the studies obtained in the denominator, one can determine if there is homogeneity of variance. That is, one can determine if all of the studies are estimating the same population effect size. If so, there are unlikely to be any moderators of the effect.

If approximately 75% of the obtained variance can be attributed to sampling error and artifacts, the effect size estimate is likely to be a homogeneous estimate of the true population effect size (also called a fixed effect). Essentially, a homogeneous effect indicates that the effect is unlikely to be moderated. In other words, the effect is unlikely to vary based on between study differences. On the other hand, if the ratio indicates that less than 75% of the variance can be so explained, the effect is likely to be heterogeneous such that there exist moderators of the effect that have not been accounted for (also called a random effect). If the effect is heterogeneous, examining the effects of moderators is useful.

To estimate the effect of such moderators on the target effect sizes, Hunter and Schmidt (2004) recommended the construction of 80% credibility intervals. These intervals use the estimate of the variance in the effect not attributable to sampling error and artifacts to provide an estimate of how widely the effect size will be likely to vary due to the moderators. For example, an effect size with a credibility interval ranging from $-0.1$ to $0.6$ suggests that there are likely moderators that will cause the effect to be reversed to a small degree (down to $-0.1$) but that most of the time, the effect will be positive (up to $0.6$). Hunter and Schmidt recommended against using confidence intervals or statistical significance tests for heterogeneity as these tend to be biased and unhelpful. Their work has also demonstrated that Fisher’s $z$ transformations tend to reduce accuracy, and consequentially, these meta-analyses used untransformed correlations.

For the continuous moderators such as year of publication or the effect sizes associated with the induction checks, metaregression techniques were used. The target effect sizes were regressed onto each hypothesized continuous moderator. The impact of each moderator on the effect sizes can then be assessed by examining the unstandardized regression coefficients. Viechtbauer’s (2010) metafor program for the R statistical package was used to calculate these regressions. That program was also used to implement the trim-and-fill algorithm to test for publication bias (Duval & Tweedie, 2000).

Duval and Tweedie’s (2000) trim-and-fill algorithm looks at the distribution of effect sizes around the average. If there are some missing effect sizes from the part of the distribution near the zero effect, the algorithm estimates how many are missing. It
then uses the existing larger effect sizes to estimate what the missing effect sizes would be. Then it recalculates the average effect with the missing studies included so that the meta-analyst can see how much smaller the average effect would be if the missing failed replications were included.

Results

The report of the various meta-analyses will begin with a description of the overall results containing all of the obtained studies. Next, the categorical moderators will be examined. Metaregression will be used to examine the effects of the continuous moderators. Finally, the publication bias tests will be reported.

Overall effects
As Table 1 shows in the first three rows, the overall results were consistent with ELM predictions. The first column shows that audiences using peripheral processing produced a small attitudinal difference between strong and weak arguments such that the strong arguments were slightly more persuasive than the weak. The second column indicates that with central processing, the strong arguments were much more persuasive than the weak, as shown by the strong effect size. The third column shows that the difference between these two effects was substantial, showing that centrally processing audiences are more affected by argument strength than peripherally processing audiences.

The fourth and fifth columns break out the separate effects for strong and weak arguments. The fourth column shows that the centrally processing audiences were somewhat more persuaded than peripheral processing audiences when exposed to strong arguments. The parallel results for the weak arguments show that central processing audiences were somewhat less persuaded by weak arguments than peripheral processing audiences (indicated by the negative effect). As these two effects are similar in size, the data are consistent with the ELM prediction that processing has the same effect regardless of whether one is focusing on the increased persuasion from strong arguments or the reduced persuasion from weak arguments.

The percentage of variance explained by sampling error and artifacts indicates that all of these effects were moderated. The pattern of these effects is largely consistent with ELM predictions, although the credibility intervals of some of these effects imply the presence of moderators that reduce the effects to nil. The potential for moderators to explain that variation will be examined next.

Categorical moderators
Aspects of the researchers
First, the field from which the study came will be examined as a potential moderator. Table 1 shows that psychology studies tended to produce slightly stronger effects than the sample as a whole and tended toward more homogeneous effects. The marketing/advertising studies tended to produce results that are a little weaker than the
### Table 1 Weighted Average Effect Sizes, Percentage of Variance Explained, and 80% Credibility Intervals for the Overall Sample and Categorical Moderators

<table>
<thead>
<tr>
<th>Field</th>
<th>Strong vs. Weak in Peripheral Processing</th>
<th>Strong vs. Weak in Central Processing</th>
<th>Difference in Argument Quality Effect C vs. P Processing</th>
<th>Strong Arguments: C vs. P</th>
<th>Weak Arguments: C vs. P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (k = 134)</td>
<td>.14</td>
<td>.47</td>
<td>.33</td>
<td>.19</td>
<td>−.2</td>
</tr>
<tr>
<td></td>
<td>.56</td>
<td>27</td>
<td>35</td>
<td>53</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>−.01, .28</td>
<td>.26, .67</td>
<td>.14, .52</td>
<td>.04, .34</td>
<td>−.39, −.01</td>
</tr>
<tr>
<td>Field</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology (k = 93)</td>
<td>.11</td>
<td>.51</td>
<td>.39</td>
<td>.22</td>
<td>−.24</td>
</tr>
<tr>
<td></td>
<td>.65</td>
<td>40</td>
<td>61</td>
<td>69</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>−.01, .24</td>
<td>.34, .66</td>
<td>.27, .51</td>
<td>.11, .34</td>
<td>−.38, −.10</td>
</tr>
<tr>
<td>Marketing/advertising</td>
<td>.17</td>
<td>.46</td>
<td>.29</td>
<td>.19</td>
<td>−.16</td>
</tr>
<tr>
<td>(k = 31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.48</td>
<td>24</td>
<td>40</td>
<td>52</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>.01, .32</td>
<td>.26, .66</td>
<td>.13, .46</td>
<td>.05, .33</td>
<td>−.41, −.10</td>
</tr>
<tr>
<td>Communication (k = 10)</td>
<td>.2</td>
<td>.24</td>
<td>.04</td>
<td>−.03</td>
<td>−.09</td>
</tr>
<tr>
<td></td>
<td>.40</td>
<td>28</td>
<td>40</td>
<td>100</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>.04, .35</td>
<td>.05, .43</td>
<td>−.12, .2</td>
<td>−.23, .05</td>
<td></td>
</tr>
<tr>
<td>Ohio State effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSU (k = 44)</td>
<td>.17</td>
<td>.55</td>
<td>.38</td>
<td>.23</td>
<td>−.24</td>
</tr>
<tr>
<td></td>
<td>.43</td>
<td>26</td>
<td>63</td>
<td>51</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>−.01, .35</td>
<td>.36, .74</td>
<td>.27, .49</td>
<td>.08, .38</td>
<td>−.45, −.05</td>
</tr>
<tr>
<td>Non-OSU (k = 91)</td>
<td>.12</td>
<td>.42</td>
<td>.3</td>
<td>.17</td>
<td>−.17</td>
</tr>
<tr>
<td></td>
<td>.67</td>
<td>34</td>
<td>30</td>
<td>57</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>−.01, .24</td>
<td>.24, .6</td>
<td>.09, .52</td>
<td>.02, .31</td>
<td>−.36, .01</td>
</tr>
</tbody>
</table>
### Table 1 Continued

<table>
<thead>
<tr>
<th>Processing induction type</th>
<th>Strong vs. Weak in Peripheral Processing</th>
<th>Strong vs. Weak in Central Processing</th>
<th>Difference in Argument Quality Effect (C vs. P)</th>
<th>Strong Arguments: C vs. P</th>
<th>Weak Arguments: C vs. P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ability (k = 13)</strong></td>
<td>Average $r$</td>
<td>.06</td>
<td>.5</td>
<td>.44</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>56</td>
<td>56</td>
<td>71</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.09, .21</td>
<td>.38, .62</td>
<td>.35, .54</td>
<td>100</td>
</tr>
<tr>
<td><strong>Motivation (k = 112)</strong></td>
<td>Average $r$</td>
<td>.15</td>
<td>.45</td>
<td>.32</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>55</td>
<td>25</td>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.01, .29</td>
<td>.26, .67</td>
<td>.12, .52</td>
<td>.02, .35</td>
</tr>
<tr>
<td><strong>Specific processing induction</strong></td>
<td>Average $r$</td>
<td>.16</td>
<td>.42</td>
<td>.26</td>
<td>.16</td>
</tr>
<tr>
<td>ORI (k = 27)</td>
<td>Percent variance</td>
<td>59</td>
<td>19</td>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.04, .28</td>
<td>.2, .65</td>
<td>.01, .51</td>
<td>−.07, .39</td>
</tr>
<tr>
<td><strong>Need for cognition (k = 6)</strong></td>
<td>Average $r$</td>
<td>.24</td>
<td>.47</td>
<td>.24</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>15</td>
<td>43</td>
<td>100</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.05, .53</td>
<td>.12, .83</td>
<td>.03, .34</td>
<td>−.53, .07</td>
</tr>
<tr>
<td><strong>Mood (k = 11)</strong></td>
<td>Average $r$</td>
<td>.07</td>
<td>.5</td>
<td>.4</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>64</td>
<td>74</td>
<td>68</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.06, .2</td>
<td>.42, .58</td>
<td>.30, .50</td>
<td>.10, .26</td>
</tr>
<tr>
<td><strong>Personality (k = 9)</strong></td>
<td>Average $r$</td>
<td>.25</td>
<td>.53</td>
<td>.27</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>79</td>
<td>24</td>
<td>90</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.17, .33</td>
<td>.33, .73</td>
<td>.22, .33</td>
<td>.14, .19</td>
</tr>
</tbody>
</table>
Table 2 Continued

<table>
<thead>
<tr>
<th>Argument quality induction</th>
<th>Strong vs. Weak in Peripheral Processing</th>
<th>Strong vs. Weak in Central Processing</th>
<th>Difference in Argument Quality Effect C vs. P Processing</th>
<th>Strong Arguments: C vs. P</th>
<th>Weak Arguments: C vs. P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thought-listing based (k = 15)</td>
<td>Average r</td>
<td>.21</td>
<td>.58</td>
<td>.37</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>27</td>
<td>18</td>
<td>100</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.03,.46</td>
<td>.37,.8</td>
<td></td>
<td>0,.36</td>
</tr>
<tr>
<td>Continuous scale based (k = 50)</td>
<td>Average r</td>
<td>.13</td>
<td>.44</td>
<td>.31</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>64</td>
<td>32</td>
<td>34</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.01,.26</td>
<td>.26,.62</td>
<td>.11,.50</td>
<td>.06,.32</td>
</tr>
<tr>
<td>Message based (k = 13)</td>
<td>Average r</td>
<td>.09</td>
<td>.31</td>
<td>.21</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>25</td>
<td>52</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.14,.33</td>
<td>.17,.44</td>
<td>−.08,.5</td>
<td>−.08,.32</td>
</tr>
<tr>
<td>Comprehensive exam messages (k = 34)</td>
<td>Average r</td>
<td>.15</td>
<td>.49</td>
<td>.34</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>100</td>
<td>32</td>
<td>46</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.31,.68</td>
<td>.18,.51</td>
<td>.07,.28</td>
<td>−.39,.05</td>
</tr>
<tr>
<td>Pre and post vs. post only</td>
<td>Average r</td>
<td>.14</td>
<td>.38</td>
<td>.25</td>
<td>.16</td>
</tr>
<tr>
<td>Pre- and posttest (k = 11)</td>
<td>Percent variance</td>
<td>89</td>
<td>100</td>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>.08,.20</td>
<td>.03,.46</td>
<td>.03,.30</td>
<td>−.40,.22</td>
</tr>
<tr>
<td>Post only (k = 123)</td>
<td>Average r</td>
<td>.13</td>
<td>.47</td>
<td>.34</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>Percent variance</td>
<td>54</td>
<td>25</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>80% Credibility</td>
<td>−.01,.28</td>
<td>.26,.68</td>
<td>.15,.53</td>
<td>.04,.34</td>
</tr>
</tbody>
</table>

Note: To calculate “Difference in Argument Quality Effect C vs. P Processing,” the effect size for argument strength under peripheral processing was subtracted from the effect size for central processing.
psychology studies. That group of studies tended toward slightly more heterogeneous effects than the psychology studies. The communication studies differed from the others quite drastically. The communication studies essentially found a main effect for argument quality such that strong arguments are somewhat more persuasive than weak arguments, regardless of processing type. The effect of switching from peripheral to central arguments among the communication studies was homogeneous, suggesting a lack of moderators. The communication studies were not consistent with ELM predictions.

The comparison between OSU researchers and non-OSU researchers will be examined next. Contrary to Johnson and Eagly’s (1989) findings, there does not seem to be a substantial OSU effect in this updated set of studies. The effect of central processing on distinguishing between strong and weak arguments was somewhat larger for OSU studies. The effects are, across the board, somewhat smaller for non-OSU studies, but all the effects remain consistently in line with ELM predictions.

Types of inductions
The comparison between ability to process inductions and motivation to process inductions shows some differences. The results in Table 1 indicate that when ability to process was low (peripheral processing), participants were unlikely to distinguish between strong and weak arguments at all. The credibility interval suggests that there are moderators that would result in low ability to process samples being slightly more persuaded by weak arguments than strong. Alternatively, for peripherally processing audiences due to low motivation rather than low ability, there were small differences between strong and weak arguments. The difference between ability and motivation to process for the central processing effects is noticeably smaller than the one found for peripheral processing.

Among the different central versus peripheral processing inductions, the results in Table 1 suggest mood as a central versus peripheral processing induction that produces the strongest effects. In these studies, participants in a positive mood generally process peripherally, whereas those in a negative mood tend to process centrally. Like those in the ability to process studies, those in positive moods seem to barely be able to distinguish strong from weak arguments and may even sometimes be more persuaded by the weak. On the other hand, those in the negative mood conditions show a strong difference between the argument types. Furthermore, the estimates of the effect sizes in these studies are all either homogeneous or nearly so. These results suggest that mood inductions were the strongest and most consistent means of affecting processing types. The other types of processing inductions did not tend to differ markedly from each other.

Next, the means of constructing argument quality inductions will be examined as a potential moderator. Table 1 shows several findings of note. First, when thought listing was used to create argument quality inductions, central processing caused a substantially larger drop in message-congruent attitudes in response to weak arguments than it did when the other types of argument quality construction
were used. In other words, when researchers used thought listing to construct strong and weak arguments, the weak arguments seem to have been particularly likely to be rejected. Next, when message qualities were altered to produce strong and weak arguments rather than pretesting them, the argument quality effects tended to be smaller. Finally, despite the justifiable concern about using the same messages in multiple studies (Wells & Windschitl, 1999), the studies that used Petty and Cacioppo’s (1986) comprehensive exam messages for strong and weak arguments did not produce effects that were substantially different than the general effects.

**Study design**

The last categorical moderator examined was whether the study’s outcome variable was a pre- and posttest message measure of attitudes or a posttest-only measure. Clearly, from the number of studies that the post-test only design is substantially more popular than the pre- and posttest design. The interaction effect shown in the third column such that central processing produced a stronger difference based on argument quality than peripheral is smaller in the pre- and posttest studies. For the weak arguments conditions, the drop in persuasiveness associated with the audience using central rather than peripheral processing was substantially smaller in the pre-post studies than the post only. The credibility interval suggests that there are moderators that would cause the argument quality effect to reverse for peripheral processors when a pre- and posttest design is used. The pre- and posttest design studies were consistent with the ELM prediction that central processing instead of peripheral processing causes strong arguments to produce more positive attitudes. Yet the results were less consistent with the prediction that central processing would produce more negative attitudes relative to peripheral processing in response to weak arguments.

**Continuous moderators**

The only continuous moderator for which information was available for all of the studies was the year of publication. Therefore, the meta-regressions could only be run with one continuous moderator variable predictor at a time rather than examining the effects of all the variables together. Table 2 shows the results of the regressions with the unstandardized regression coefficients and their associated p-values. The weighted average estimates of the continuous predictors will each be reported before assessing their impact on the various effect size outcomes.

**Year of publication**

First, the year of publication was examined as a potential moderator. As Table 2 shows, for several of the predicted effect sizes, year of publication had a statistically significant negative effect. Essentially, for the difference in the effect of argument quality between central and peripheral processing, every 10 years has brought a decrease in the difference of that effect of about .05.²
Table 2 Results of the Metaregression Analysis of Continuous Moderators

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Strong vs. Weak in Peripheral Processing</th>
<th>Strong vs. Weak in Central Processing</th>
<th>Difference in Argument Quality Effect C vs. P Processing</th>
<th>Strong Arguments: C vs. P</th>
<th>Weak Arguments: C vs. P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>−.002</td>
<td>−.007</td>
<td>−.005</td>
<td>−.004</td>
<td>.004</td>
</tr>
<tr>
<td>p</td>
<td>.34</td>
<td>.0002</td>
<td>.01</td>
<td>.04</td>
<td>.06</td>
</tr>
<tr>
<td>Processing induction check</td>
<td>.24</td>
<td>.19</td>
<td>.02</td>
<td>.14</td>
<td>.19</td>
</tr>
<tr>
<td>p</td>
<td>.06</td>
<td>.24</td>
<td>.99</td>
<td>.43</td>
<td>.35</td>
</tr>
<tr>
<td>Argument quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pilot rest</td>
<td>−.27</td>
<td>−.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.15</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>induction check</td>
<td>.23</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.1</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $b =$ unstandardized regression coefficient.

Processing induction checks

The second potential moderator was the size of the correlation for the induction check in the studies that estimated the difference in processing depth between the central and the peripheral conditions. There were 44 studies that provided enough information to calculate the strength of the processing type induction check. The weighted average effect size for the induction checks was average $r = .34$, 80% CredI [.15, .53] with 39% of the variance explained by sampling error.

The induction checks for processing depth were predicted to have their largest effect on the difference between the effects of argument quality on central versus peripheral processing, such that larger induction checks are positively associated with a larger difference. Surprisingly, the unstandardized regression coefficient was nearly zero. This finding indicates that induction checks are unlikely to be associated with the key prediction of the ELM.

The finding that communication studies tend to fail to replicate the ELM may be the result of methodological problems. With that possibility in mind, the same metaregressions were conducted on the sets of studies from each field (psychology, marketing/advertising, and communication). The only instances where there was a deviation from the general pattern that was theory consistent were the
results associated with the communication studies. When the difference in the argument strength effect between central and peripheral processing audiences was regressed on the processing type induction check, the result was substantial and statistically significant, \( b = .50, p = .04, k = 5 \). Essentially, the key result expected to be associated with the ELM was associated with its induction check, but only among communication studies. In addition, the theory would also predict that a large induction check to show greater central processing would be associated with a stronger argument strength for a centrally processing audience. This predicted result was also found only among the communication studies, \( b = .49, p = .04, k = 5 \). The relationship between the processing induction check and the argument strength effect among the peripherally processing audiences was trivial in size and not statistically significant. These findings need to be accepted tentatively as there were only five communication studies with a reported processing induction check. There was not any substantial variation in the metaregression results for any of the other tests based on the field in which the study was conducted.

**Argument quality pilot tests**

Next, the effect of pilot tests of the argument quality inductions will be examined. Of the studies obtained, 27 described a test of the argument-quality induction with a separate sample with sufficient detail reported to calculate an effect size for the difference between the strong and the weak arguments. The weighted average effect size for the argument-quality pilot tests was average \( r = .56, 80\% \text{ CredI} [.34, .79] \) with 23% of the variance explained by sampling error. The low explained variance and wide credibility interval suggest a great deal of variation in the strength of the argument quality inductions.

The pilot tests for argument quality were expected to have a substantial and positive impact on the effect size for strong versus weak arguments in centrally processing audiences but have a substantially smaller effect on peripherally processing audiences. Neither effect was statistically significant.

**Argument quality induction checks**

Next, the induction checks for argument quality will be examined. These represent the impressions of the quality of the arguments reported by participants in the primary studies. There were 58 studies that reported an argument quality induction check with sufficient detail to calculate an effect size. The weighted average effect size for the argument quality induction checks was average \( r = .40, 80\% \text{ CredI} [.23, .56] \) with 41% of the variance explained by sampling error.

The induction checks for argument quality were expected to have the same effects as those of the argument quality pilot tests. The unstandardized regression coefficients were insubstantial and not statistically significant. In addition, the effects of the argument quality induction checks were nearly equal regardless of whether the audience was processing peripherally or centrally.
Publication bias
To determine if the estimates are biased by missing studies due to a bias against publishing null findings, the trim-and-fill algorithm was employed (Duval & Tweedie, 2000; Viechtbauer, 2010). The primary place in which one would expect a publication bias is for studies that failed to find the interaction between processing type and argument quality. Such an interaction is represented by the difference in the effect size of argument quality in the central and peripheral conditions. The trim-and-fill analysis estimated that there are 31 studies missing from the dataset. That estimate indicates that approximately 19% of the studies conducted on this interaction are either not published or otherwise missing from the dataset.

The trim-and-fill algorithm also uses the information in the dataset to estimate how the effect size estimate would change if the missing studies were present. Results indicated that the size of the difference in the correlations would drop by about .06. Although these results do indicate a slight overestimate in the effect size, they are still consistent with the ELM’s prediction that the effect of argument quality would be larger when the audience is processing centrally than peripherally. This drop in effect size is consistent with Carpenter’s (2012) finding that the overestimation in effect sizes among meta-analyses of communication topics tends to be below .1, when they occur at all. This estimate should be interpreted with caution, however, as the trim-and-fill algorithm may overestimate publication bias in the presence of heterogeneity (Peters, Sutton, Jones, Abrams, & Rushton, 2007).

The trim-and-fill algorithm can also be used to test Johnson and Eagly’s (1989) speculation that failed replications of dual-process studies conducted at OSU would be less likely to be submitted for publication than studies conducted elsewhere. Again the focus was on the difference in the effect of argument quality under central versus peripheral processing. When the studies not conducted at OSU were considered as a set, the trim-and-fill algorithm failed to detect publication bias. When the OSU studies were considered as a set, it again failed to detect any publication bias. These results are not consistent with the prediction of greater publication bias by OSU scholars.

Discussion
These meta-analyses were conducted to estimate the effects associated with the ELM and HSM prediction that argument quality would have a diminished effect on attitudes when the audience was processing peripherally rather than centrally. In addition, several moderators suggested by critics of the theory and the diversity of inductions were examined. The discussion of the findings will examine those findings that were consistent with the ELM, the findings inconsistent with the ELM, those that indicate interesting variations, and then those that suggest methodological problems with the way the ELM has been tested.

Findings consistent with the ELM
The key prediction derived from the ELM was that there would be a difference in the size of the argument quality effect on attitudes such that the effect would be
substantially larger when the audience was processing centrally than when it was processing peripherally. The estimate of the weighted average effect was consistent with that prediction and the credibility interval indicated that although moderators may shrink that difference, it is likely to remain consistent with the theory. When the effect of central compared to peripheral processing was focused on strong arguments, central processors were consistently more likely to be persuaded than peripheral, although the effect was not large. A parallel effect was found such that when the arguments were weak, central processors were consistently less likely to be persuaded than peripheral.

A previous meta-analysis that focused on involvement found an “Ohio State effect” such that OSU researchers found the predicted interaction between involvement and argument quality, whereas non-OSU researchers did not (Johnson & Eagly, 1989). The current meta-analysis, which included the many studies published since then and included many processing type moderators, found a much diminished OSU effect. The effects were consistently a little larger among the OSU studies, but most of the non-OSU studies found the predicted interaction. In addition, the test for publication bias did not find evidence of OSU researchers being less likely to publish null results than non-OSU researchers.

Findings not consistent with the ELM
Several findings were, on the other hand, inconsistent with ELM predictions. First, the data suggested that the effects are getting smaller over time. This finding raises the possibility that as other labs attempt to replicate the ELM after the initial strong demonstrations, the results become less consistent. Another possibility is that as the theory has diffused, researchers have been applying it in an ever-widening variety of contexts, some of which may be beyond the boundary conditions of the theory. Researchers may be trying a wider variety of means of inducing the key constructs of central versus peripheral processing and argument quality, and some of these may be less successful. This finding is, however, consistent with the growing reports of “decline effects” in which published effects in a research area get smaller over time (Schooler, 2011).

Perhaps more important than the decreasing effect sizes, however, was the substantial differences between the studies that employed a pre- and posttest design to measure attitude change instead of the more common posttest-only design. The pre-post designs found consistently smaller and in some cases null effects. This difference was particularly acute for the reduction in persuasiveness that was meant to accompany weak arguments when central processing was used instead of peripheral.

This finding is consistent with Hamilton et al.’s (1993) critique that the ELM studies show attitude formation rather than change. The finding was also consistent with their assertion that weak arguments are unlikely to produce a boomerang effect, defined as a message produces attitudes that are less in favor of the message position after hearing it. The pre-post studies suggested that central processing may cause someone to respond to strong arguments by moving closer to the message position
more than peripheral processing. Yet for weak arguments, when actual attitude change was measured, participants who were centrally processing were often no more likely to move further from the message position than peripheral processors. There were only 11 studies with a pre- and posttest design so these findings must be considered tentative until further studies can be conducted.

Although the OSU effect seems to have diminished, a communication researcher effect seems to have appeared. Psychology studies produced somewhat larger effects than studies promoting consumer products, but communication researchers seemed particularly unlikely to find data consistent with the ELM. One possible explanation for this finding is that communication researchers are more likely to focus on message aspects of argument quality (O’Keefe, 2003).

Message-based inductions of argument quality were more common than the other types of argument quality inductions among the communication studies. Johnson, Smith-McLallen, Killeya, and Levin (2004) found that argument quality inductions that induced more or less desirable consequences tended to interact with central versus peripheral processing as expected by the ELM, whereas argument quality inductions based on likelihood of the outcomes did not interact with processing type. One prominent failed ELM replication from communication researchers (Park et al., 2007) induced argument quality by strengthening the logic of the arguments—an induction that seems more likely to affect the likelihood of the argument being true than its desirability. One possibility is that the amount of cognitive processing required to detect the desirability of the outcomes described in the message was lower than the amount required to assess their likelihood.

Investigations exploring these effects may consider several means of testing these possibilities. One option is to employ various processing type inductions. Then the researchers might separately measure the extent to which an audience processes outcome desirability and the extent to which they process outcome likelihood. Self-report measures of these different processes would need to be developed to pursue such inquiries. The ELM has argued that there is a single continuum of processing (Petty & Cacioppo, 1986) but perhaps the various cognitive activities that the ELM combines may be more distinct than they predicted.

Interesting variations
The possibility was explored that the different means of inducing central versus peripheral processing might affect the size of the effects. Specifically, variations in processing inductions might alter how much larger the argument quality effect is for central than for peripheral processing. The analysis indicated that studies employing ability to process inductions as a class (distractions, confusing text, etc.) produced stronger effects than those associated with changing levels of processing motivation. In particular, when ability to process the message was low (peripheral processing), the participants were particularly unlikely to distinguish between strong and weak arguments. Motivation can be induced to be higher or lower, but there may be plenty of people who decide to process more than their induction would motivate them to
(e.g., those high in need for cognition; Cacioppo & Petty, 1982). An induction that prevents participants from processing may be more likely to prevent even those high in need for cognition from processing centrally.

In addition, the mood inductions produced particularly strong effects. In these studies, people in negative moods (usually sad) process centrally, whereas those in positive moods (usually happy) process peripherally (Bless, Mackie, & Schwarz, 1992; Worth & Mackie, 1987). The difference in the size of the effect of argument quality between the centrally and the peripherally processing audiences was among the largest of the subgroups examined. The effect was particularly large for the effect of central processors being more likely to reject weak arguments than peripheral processors.

Interpretation of these variations is, however, difficult. Perhaps these different inductions are not all inducing a single construct of “processing type,” or perhaps they are inducing different points along the elaboration continuum. Unfortunately, there is no consistently used means of assessing precisely how much the audience is processing. The induction checks are always reported relative to the other processing conditions. The condition that is induced as “peripheral” processing in a given study may show a statistically significant difference in the processing induction check from the “central” processing condition. Yet that information only tells the researcher how far apart on the elaboration continuum the two conditions are from each other rather than where they are on an absolute standard. It does not tell the researcher if the peripheral processing condition is in fact at a low point on the continuum or just lower than the central condition. Only future work that attempts to precisely estimate how much the audience is processing can determine if the stronger effects associated with some processing inductions are due to inducing different constructs or simply stronger effects of the same construct.

**Methodological problems**

These meta-analyses also revealed some problems with the two key variables of interest. First, both inducing and assessing argument quality remain problematic. Petty and Cacioppo’s (1986) initial description of the ELM made it clear that they were uninterested in what message variables produce argument quality. In subsequent reviews, they continued to state that argument quality is only defined empirically such that strong arguments produce positive thoughts and weak arguments negative ones (Petty, Rucker, Bizer, & Cacioppo, 2004). This analysis found that when researchers stick to either the ELM preferred method of using thought-listing pilot tests or the alternate method of using rating scales for pilot tests, the effects are similar. These methods both sidestep the issue of what causes arguments to be strong or weak.

In contrast, when researchers have attempted to induce argument quality by focusing on the content of the message, the ELM effects tend to be diminished. Consistent with the difficulty in conceptualizing what is being induced, Reimer, Mata, and Stoecklin (2004) found that argument quality inductions can affect the perceived credibility of the speaker as well as perceptions of argument quality, confounding different persuasion variables. Clearly, more research is needed to determine what message
features interact with central versus peripheral processing as argument cues. Although social psychologists such as Petty and Cacioppo may remain indifferent, the field of communication seems particularly well suited to attempt to determine what message features do interact with processing and produce stronger arguments (O’Keefe, 2012).

In addition to difficulties in inducing argument quality, this analysis revealed problems with measuring it. The most surprising finding concerned argument quality pilot tests that were conducted with a separate sample to establish an argument quality induction. Many of the studies that used the ELM’s thought listing method of pilot testing argument quality could not contribute to the tests of this effect in the meta-analyses as those studies do not produce precise estimates of that test. Such studies usually just report that the thoughts for the strong arguments were mostly positive and for the weak, mostly negative. These reports make it impossible to calculate exactly how strong the argument quality induction was for those studies.

For the studies that did report enough data about their pilot test to use the estimate as a moderator, the result was that the argument quality induction strength was unlikely to be related at all to the effect of argument quality on attitudes in the main study. The effect of argument quality induction checks in the primary study were not much better as argument quality induction checks were not substantially associated with increases in argument quality effects on attitudes. Perhaps the measures used to assess argument quality are lacking as they are often ad-hoc and single-item measures. A new measure has been developed and may improve the ability of researchers to assess the quality of the arguments used in such studies (Zhao, Strasser, Cappella, Lerman, & Fishbein, 2011).

Another possibility is that argument quality is not a unitary construct. Morley (1987) argued that there were three aspects to argument quality that are conceptually distinct. That article presented evidence that variations in the persuasiveness of messages can be explained by the three constructs rather than a single rating. Simply measuring argument quality as a single construct may be obscuring the differences between variables that produce changes in perceived argument quality but do not affect the persuasiveness of the message. Although argument quality may have been a convenient methodological tool for ELM researchers, whether it is a single construct that varies on a single continuum remains to be seen.

The other key variable in the dual-process theory is central versus peripheral processing. The exact nature of these two types of processing has been the subject of much debate (Hamilton et al., 1993; Kruglanski & Thompson, 1999). Petty and Cacioppo (1986) described central processing as “the process of relating the to-be-evaluated recommendation and other arguments to other issue-relevant information in memory” (p. 14). They also stated that elaboration involves, “the process of vigilantly scrutinizing the arguments and information in memory bearing specifically on the desirability of a persuasive recommendation” (p. 15).

They conceived of these cognitive activities as those that can be done more or less and referred to them as the elaboration continuum. They also explained that although high amounts of cognitive effort is usually required for such processing, effort is an
orthogonal construct. Petty and Cacioppo reviewed several methods of assessing the degree to which an audience is processing centrally and concluded that the best indicator is the extent to which argument quality affects the ratio of positive to negative thoughts. Rather than use an indicator of central processing, they point to the size of the relationship across participants. Researchers are left without a way to determine which participants are centrally processing and which are peripherally processing.

Their skepticism about other measures of central versus peripheral processing may have been justified based on the findings of these meta-analyses. The induction checks for central versus peripheral processing had almost no effect on the extent to which the argument quality effect was larger for central processing relative to peripheral processing. This finding suggests that the either the ELM is invalid or the current methods of assessing central and peripheral processing in each subject are severely lacking. If it is the latter, persuasion researchers must develop a better measure than the current collection of ad-hoc and single-item measures employed in these studies. As communication researchers were the only group for whom such induction checks were associated with the predicted effect sizes, their methods may be useful.

There are two validated measures of outcome-relevant involvement (Cho & Boster, 2005; Miller & Averbeck; 2013). There is also the need for a cognition measure (Cacioppo & Petty, 1982). These, unfortunately, only cover two of the many hypothesized causes of processing type. What is required is a general measure. Reynolds (1997) published such a measure, but very few researchers have tested it in an ELM context. Furthermore, Carpenter and Boster (2013) found that the Reynold’s measure had two factors that operated differently. Additional scale development work is required to either produce a new measure or revise Reynolds’s.

In addition, the average induction check for processing type produced a medium-sized effect ($\rho = .34$). If the average study has such a moderately sized induction, it suggests that the other relationships dependent on that induction will be weaker still. For example, when the audience is exposed to strong arguments, the ELM would predict that the extent to which they are processing centrally will be positively associated with message-congruent attitudes. If the average induction check is moderate in size, the induction itself will necessarily have an even weaker effect on attitudes, if the model is correct. Furthermore, the range of the effect size for the induction check was from .13 to .85. Generally, researchers will assess whether their induction check is statistically significant and ignore how large the effect is. Researchers conducting persuasion research would do well to carefully examine how big of an effect their induction is having rather than simply relying on statistical significance (O’Keefe, 2003).

**Conclusion**

This study set out to assess one of the predictions of the ELM by meta-analyzing the studies that tested the Central versus Peripheral $\times$ Weak versus Strong Arguments interaction. Although the data are largely consistent with the ELM predictions (for
the post-test only studies), they were not consistent with the ELM explanation, given the failure of the induction checks of the hypothesized explanatory variables to consistently predict the size of the outcomes. Future research requires greater methodological sophistication if the processes that underlie these findings are to be understood.

Acknowledgments

The author would like to thank Mike Kotowski and Dan O’Keefe for supplying additional articles.

Note

1 Effect sizes were calculated by using available information to calculate a correlation coefficient. Specifically, the means, standard deviations, and sample sizes were converted into correlation coefficients. When standard deviations were not provided, they were estimated from between-subjects ANOVAs. See Lipsey and Wilson (2001) for formulas.

2 This estimate was calculated by using the unstandardized regression coefficient to construct the linear model $r = -0.005(years)$ and then inserting 10 for the years $r = -0.005(10)$ to produce the estimate that $r$ would decline by .05 every 10 years after the first publication of a dual-process study.

References


**Appendix**


