Do incompatible arguments cause extensive processing in the evaluation of arguments? The role of congruence between argument compatibility and argument quality

Cheng-Hong Liu1*, Hung-Wei Lee2, Po-Sheng Huang2, Hsueh-Chih Chen3 and Scott Sommers4

1Center for Teacher Education, National Tsing Hua University, Hsinchu City, Taiwan
2Department of Applied Psychology, Hsuan Chuang University, Hsinchu City, Taiwan
3Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei City, Taiwan
4English Learning Center, Ming Chuan University, Taipei City, Taiwan

Previous studies have demonstrated that arguments incompatible with prior beliefs are subjected to more extensive refutational processing, scrutinized longer, and judged to be weaker than arguments compatible with prior beliefs. However, this study suggests whether extensive processing is implemented when evaluating arguments is not decided by argument compatibility, but by congruence between two evaluating tendencies elicited by both argument compatibility and argument quality. Consistent with this perspective, the results of two experiments show that relative to congruent arguments, participants judged arguments eliciting incongruent evaluating tendencies as less extreme in strength, spent more time, and felt more hesitant generating strength judgments for them. The results also show that it is mainly incongruent arguments, not congruent arguments, whose strength ratings were more closely associated with the perceived personal importance of the issue, which intensified the tendency to evaluate arguments depending on argument compatibility. These results suggest that it is the incongruity between argument compatibility and argument quality, and not simply the argument compatibility, that plays a more important role in activating an extensive processing in the evaluation of arguments.

Are you for or against the death penalty? Do you support the legitimacy of abortion? On a daily basis, we encounter such issues and a wide range of conflicting opinions from the people around us. The evaluation of received arguments is very important. It influences our attitudes on related issues (Clark, Wegener, & Fabrigar, 2008; Lord, Ross, & Lepper, 1979) and even our decisions and actions (e.g., Ditto & Lopez, 1992). However, previous studies indicate that people are not always rational when evaluating arguments (e.g., Bastardi, Uhlmann, & Ross, 2011; Fischer, Greitemeyer, & Frey, 2008; Stanovich & West, 1997, 1998). A clarification of the processes through which people evaluate arguments is of great importance, not only for advancing our conceptual

*Correspondence should be addressed to Cheng-Hong Liu, 101, Sec. 2., Kuang Fu Road, Hsinchu City 30013, Taiwan (email: liusmallhon@seed.net.tw).
understanding, but also because it provides guidance for improving rational sharing of opinions with others.

In the past, researchers have focused on examining the relationship and processes operating between prior beliefs and the evaluation of received evidence and arguments. Lord et al. (1979) found that participants evaluated the arguments compatible to their prior beliefs as stronger compared with incompatible arguments. Ditto and Lopez (1992) found that when participants were provided with unfavourable medical test results, they rated the test as less accurate compared to participants who received a favourable diagnosis. Edwards and Smith (1996) suggested the term prior belief effect to describe the phenomena and proposed the disconfirmation model to explain the underlying processes. This model argues that when a person is presented with an argument to evaluate, there will be some automatic activation in memory of material relevant to the argument, inclusive of the prior beliefs about the issue. If the argument is compatible with prior beliefs, it is evaluated quickly. If, however, the argument is incompatible with prior beliefs, there will be a deliberate search for evidence to undermine the argument. In brief, incompatible arguments would cause more extensive refutational analyses and thus would be judged as weaker.

Several researchers have also proposed similar descriptions. For example, Ask and Granhag (2007) asked experienced criminal investigators to evaluate a witness’s testimony for a homicide case. The results showed that testimonies which disconfirmed the investigators’ hypothesis were perceived as less reliable and credible. In addition, some researchers emphasized that people tended to rely on the face value of an argument to evaluate compatible arguments while spending more time scrutinizing incompatible ones (e.g., Chaiken, Liberman, & Eagly, 1989; Koehler, 1993; Ross & Lepper, 1980). A number of other researchers have also extended these ideas to explore issues related to biased information processing, such as persuasion (e.g., Clark et al., 2008), memory (e.g., Eagly, Kulesa, Chen, & Chaiken, 2001), and motivated political reasoning (Taber, Cann, & Kucsova, 2009). However, is the compatibility of arguments with one’s prior beliefs (argument compatibility) the only key factor causing the extensive processing in the evaluation of arguments? There is reason to believe other attributes, such as argument quality should also be considered (Stanovich & West, 1997, 1998; Wood, Kallgren, & Preisler, 1985).

In fact, when we are evaluating the strength of an argument, it is not only the biased tendency elicited by the argument compatibility that influences our judgments but also the objective quality of the argument that plays an important role. A strong argument, which has an evident relationship between the premise and the conclusion and has content that is objectively regarded as an important aspect of the issue, tends to be rated as more positive. However, a weak argument, which has only a vague causal relation between its premise and conclusion or contains content which is objectively regarded as less important to the issue, is usually evaluated as more negative (Crites, 1965; Edwards & Smith, 1996; Ranganath, Spellman, & Joy-Gaba, 2010; Smith, 1977; Stanovich & West, 1997, 1998). This is consistent with research findings on attitudes. Although not all researchers are in agreement with each other about the operational definition of argument quality, studies support the idea that strong arguments are positively and weak arguments are negatively associated with attitude judgments (e.g., Bohner, Erb, & Siebler, 2008; Van Overwalle & Siebler, 2005).

If both argument compatibility and argument quality play important roles in evaluating arguments, this would mean people could simultaneously hold two different evaluating tendencies towards the same argument and this would produce two independent main
effects on argument evaluation (Bassili, 2008; Eagly & Chaiken, 2007; Stanovich & West, 1997, 1998). In such a case, do the processes specified in the disconfirmation model reflect the processes underlying argument evaluation? Because Edwards and Smith (1996) aimed at identifying the mechanisms underlying the prior belief effect, they did not include an argument’s quality as a variable in the disconfirmation model. Also, they used only moderately strong arguments as experimental materials to confirm their predictions. Thus, it is possible that the disconfirmation model is only valid for arguments possessing moderately strong quality. For this reason, we consider both the influences of argument compatibility and argument quality and try to clarify the processes underlying argument evaluation.

The congruence model
We propose a new model called the congruence model (see Figure 1) to specify the processes underlying argument evaluation. The model makes several key assumptions.

1. When people are evaluating an argument, relevant material is automatically activated in memory. This activated portion of cognition includes prior beliefs about the issue (Chaiken et al., 1989; Edwards & Smith, 1996; Petty & Cacioppo, 1986), as well as some rules for judging whether the argument is reasonable (Kunda, 1990).

2. Both the compatibility and the quality of an argument will, respectively, elicit positive or negative tendencies during evaluation. First, if the argument is compatible with people’s prior beliefs, it tends to cause a positive evaluating tendency. If the argument is incompatible, there is a negative evaluating tendency. This is called the compatibility tendency. In addition, if the argument is a strong argument, it tends to cause a positive evaluating tendency because it conforms to the rules that judge an argument as reasonable (Bohner et al., 2008; Kunda, 1990). If the argument is weak, there is a negative evaluating tendency. This is termed the quality tendency. In brief, these two evaluating tendencies are assumed to simultaneously affect argument evaluation (Bassili, 2008; Eagly & Chaiken, 2007; Stanovich & West, 1997).

3. When these two evaluating tendencies are congruent in valence (i.e., evaluating a compatible and strong argument or an incompatible and weak argument), there is no intrapsychic conflict. For the sake of limited cognitive resources, the economically minded person needs no extra processing at all (Chen & Chaiken, 1999; Kunda, 1999) and would just follow the congruent tendencies to evaluate the arguments as more extreme. Specifically, when evaluating compatible–strong arguments, people are motivated to rate these arguments as stronger because these arguments are...
‘compatible’. They also know that they should rate these arguments as stronger because these arguments are ‘strong’. Thus, no additional effort is required to reconcile these tendencies and people would quickly rate these arguments as most positive. On the contrary, when evaluating incompatible–weak arguments, people are motivated to rate these arguments as weaker because these arguments are ‘incompatible’. They also know that they should rate these arguments as weaker because these arguments are ‘weak’. Thus, these arguments would be quickly rated as most negative.

4. When these two evaluating tendencies are incongruent in valence (i.e., evaluating a compatible but weak argument or an incompatible but strong argument), this incongruity in a message would result in intrapsychic conflict which can lead to enhanced information processing (Baker & Petty, 1994; Maheswaran & Chaiken, 1991). That is, intrapsychic conflict would drive them to extensively process the argument (Clark, Wegener, Habashi, & Evans, 2012; Clark et al., 2008; Festinger, 1957; Newby-Clark, McGregor, & Zanna, 2002) and reconcile between two contradicting tendencies to reach a more moderate conclusion (Bohner et al., 2008; Simon, Snow, & Read, 2004; Van Overwalle & Siebler, 2005). After such a time, people will evaluate these arguments as less extreme in strength (Kunda, 1990). Specifically, when evaluating compatible–weak arguments, people are motivated to rate them as stronger because these arguments are ‘compatible’. At the same time, people may also know these arguments should be rated as weaker because these arguments are ‘weak’. Thus, the contradicting evaluating tendencies would drive people to perform extensive reconciliation processing which would cause people to rate them as more moderate. Similarly, people are motivated to rate incompatible–strong arguments as weaker because these arguments are ‘incompatible’. People may also know these arguments should be rated as stronger because these arguments are ‘strong’. Thus, after the extensive reconciliation processing, people may evaluate these arguments as more moderate.

The present research

The congruence model emphasizes that argument compatibility is not the only factor influencing whether extensive processing is implemented when evaluating arguments. Incongruity between evaluating tendencies from argument compatibility and argument quality plays a more important role in activating an extensive reconciliation processing. To clarify these, participants in two studies rated the strength of congruent (i.e., compatible–strong and incompatible–weak) and incongruent (i.e., compatible–weak and incompatible–strong) arguments. In study 1, we aimed to examine whether participants judge incongruent arguments as more neutral in strength and spend more time generating judgments for them compared with congruent arguments. In study 2, in addition to further examining whether participants feel more hesitant generating judgments for incongruent arguments, we also explored the effects of participants’ perceived personal importance of the issue on the strength ratings of different types of arguments.

EXPERIMENT 1

We asked participants to make computerized evaluations of the strength of arguments about abortion. First, according to the congruence model that compatible–strong and
incompatible–weak arguments elicit congruent evaluating tendencies, we predicted that the former will be evaluated as the strongest and the latter as the weakest arguments. However, because compatible–weak and incompatible–strong arguments elicit conflicting tendencies that could be further reconciled, they will be evaluated as more neutral in strength. Next, because the individual has to spend extra time to reconcile conflicting tendencies, we predicted that incongruent arguments will take longer to evaluate compared with congruent ones. These predictions are counter to those made from the disconfirmation model. Because the disconfirmation model argued that incompatible arguments drive individuals to implement refutational analyses which are time-consuming, incompatible (incompatible–strong and incompatible–weak) arguments should take longer to be evaluated and be rated as weaker than compatible (compatible–strong and compatible–weak) ones.

Method

Participants and design
A total of 40 undergraduates at a university in western Taiwan received payment and extra course credit for participation. The positions (pro- vs. anti-) and qualities (strong vs. weak) of arguments about the legitimacy of abortion were manipulated, and the participants’ existing positions on this issue were measured. The ratings of argument strength and evaluating time were dependent variables.

Materials
We first drafted 14 arguments supporting and 13 arguments opposing the legitimacy of abortion. About half of the arguments were strong and the others were weak. Each argument was expressed in a single sentence which clearly indicated the advocated position of the argument (this is similar to the conclusion of the argument) and a reason to support the advocated position (this is similar to the premise of the argument). In a strong argument, the causal relationship between the reason and the advocated position is more evident and the content is an important aspect of the issue in our cultural context. A weak argument possesses only vague causal relationship, or the content is not an important aspect. Combining argument position (pro- vs. anti-) and quality (strong vs. weak), four types of arguments were generated, such as ‘legalizing abortion can reduce social problems caused by the birth of illegitimate children’ (a pro-strong argument), ‘legalizing abortion can reduce some unnecessary medical costs’ (a pro-weak argument), ‘opposition to abortion is necessary for us to protect the rights of the unborn’ (an anti-strong argument), and ‘opposition to abortion can prevent declining birth rate in the future’ (an anti-weak argument).

Similar to Stanovich and West (1997), we used a summary measure of experts’ evaluations to assure differentiation of arguments’ strength in our cultural context. We invited five experts (full-time professors in philosophy teaching logic at universities in Taiwan) and one of the authors to rate each argument’s strength on a 6-point scale (1 = very weak, 6 = very strong). We chose 16 arguments which were most representative of each of the four types of arguments (four arguments for each type). The median correlation between the judgments of the six experts was .58. If the author’s ratings were excluded, the median correlation between the five experts was .52 (ps < .05). Expert opinion was reasonably consistent for ratings of argument quality. The mean strength ratings for pro-strong, pro-weak, anti-strong, and anti-weak arguments were 5.21, 2.92,
5.04, and 2.71, respectively. A repeated-measures ANOVA revealed the main effect of argument type on the strength ratings, $F(3, 15) = 18.80, p < .001, \eta^2 = .79$. The LSD post-hoc test showed that there were no differences between pro-strong and anti-strong arguments ($p = .516$) or between pro-weak and anti-weak arguments ($p = .341$). However, pro-strong and anti-strong arguments were both rated as stronger than pro-weak ($ps = .006, .003$) and anti-weak arguments ($ps = .015, .005$).

Next, we recruited 78 undergraduates from two universities in northern Taiwan to rate the strength for each argument. This pilot study was disguised as a ‘Test of Critical Thinking’ and participants were told to rate the argument strength as accurately as possible. A repeated-measures two-way ANOVA revealed that there was no main effect for argument position, $F(1, 77) = 0.81, p = .370$, nor did argument position interact with argument quality, $F(1, 77) < 0.01, p = .983$. This indicated that the arguments for different positions were approximately equal in strength. In addition, strong arguments were viewed as stronger ($M = 4.57, SD = 0.72$) than weak arguments ($M = 2.85, SD = 0.76$), $F(1, 77) = 339.47, p < .001, \eta^2 = .82$.

Procedures and measurement instruments
Five to eight people arrived for each individual session and were seated separately at a computer station. In stage 1, participants were introduced to the fictitious purpose of the study (a survey on the issue of the legitimacy of abortion) and instructed to indicate their positions on this issue using a two-alternative forced choice (2AFC) on the first page of a packet. In stage 2, participants read messages displayed on monitors. The instructions introduced the participants to a fictitious undergraduate whose arguments on both sides of this issue they were to evaluate. Participants were reminded to make their judgments without hesitation. Finally, participants rated the strength of 16 arguments presented in a random order. Each argument was presented along with a 6-point scale at the bottom of the screen ranging from very weak (1) to very strong (6). Time elapsed during evaluation of each argument was recorded in milliseconds. In stage 3, participants retrieved the packet and continued to page 2. In addition to some filler items, participants answered one question addressing ‘the precise position you possess on the issue after rating all the arguments’ (1 = extremely anti-abortion, 6 = extremely pro-abortion). Finally, they were debriefed and thanked for their participation.

Results
Preliminary analyses
Both the pre-judgment position in stage 1 (2AFC, $M = 1.70, SD = 0.46$) and post-judgment position in stage 3 (6-point scale, $M = 4.20, SD = 1.22$) of the 40 participants were generally favourable towards pro-abortion. The correlation between the pre- and post-judgment positions was .61 ($p < .001, N = 40$). There were two participants who did not keep the same position in the pre- and post-judgment stage. After excluding their data, the correlation between the pre- and post-judgment positions was .83 ($p < .001, N = 38$).

Test of predictions
We first created an argument compatibility variable. When the position advocated in an argument is consistent with the participant’s position, it is categorized as a compatible
argument. If inconsistent, it is an incompatible argument. Then, a repeated-measures two-way ANOVA conducted on the strength ratings showed that the interaction of argument compatibility and argument quality was not significant, $F(1, 37) = 0.71, p = .407$ (see Table 1). However, compatible arguments were rated as stronger ($M = 4.25, SD = 0.58$) than incompatible ones ($M = 3.86, SD = 0.65$). Strong arguments were stronger ($M = 4.77, SD = 0.63$) than weak ones ($M = 3.34, SD = 0.67$), $F_{s}(1, 37) = 20.03, 138.18, ps < .001, \eta^2 = .35, .79$. The results indicate that argument compatibility and argument quality elicit independent evaluating tendencies. A LSD post-hoc test showed that compatible–strong arguments ($M = 5.01, SD = 0.69$) were judged as strongest and sequentially followed by incompatible–strong ($M = 4.53, SD = 0.73$), compatible–weak ($M = 3.50, SD = 0.79$), and incompatible–weak ones ($M = 3.18, SD = 0.83$).

Next, considering evaluating time, we excluded a further two participants’ data because their evaluating time on more than half of arguments (for each type) exceeded three standard deviations from the mean. The results revealed no main effects of argument compatibility and argument quality, $Fs < 1.38, ps > .05$. However, the interaction was significant, $F(1, 35) = 10.01, p = .003, \eta^2 = .22$ (see Figure 2). Simple effects analysis revealed that evaluating time for compatible–weak arguments ($M = 6,872, SD = 2,134$) tended to be longer than incompatible–weak ones ($M = 6,287, SD = 1,727$), $F(1, 70) = 3.22, p = .077, \eta^2 = .04$. And evaluating time for incompatible–strong arguments ($M = 7,196, SD = 2,082$) was longer than compatible–strong ones ($M = 6,399, SD = 1,836$), $F(1, 70) = 5.96, p = .017, \eta^2 = .08$.

**Mediation analysis**

The results showed that participants judged incongruent arguments as less extreme in strength and tended to spend more time generating judgments for them. However, if the participants evaluated incongruent arguments as less extreme primarily because these arguments caused a deliberative reconciliation processing, the evaluating time should mediate the effect of argument (valance) congruency on the strength ratings. To test this

<table>
<thead>
<tr>
<th>Argument compatibility</th>
<th>Strong argument</th>
<th>Weak argument</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strength ratings</td>
<td>Evaluating time (ms)</td>
<td>Strength ratings</td>
</tr>
<tr>
<td>Compatible argument</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>5.01</td>
<td>6,399</td>
<td>3.50</td>
</tr>
<tr>
<td>SD</td>
<td>0.69</td>
<td>1,836</td>
<td>0.79</td>
</tr>
<tr>
<td>Incompatible argument</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4.53</td>
<td>7,196</td>
<td>3.18</td>
</tr>
<tr>
<td>SD</td>
<td>0.73</td>
<td>2,082</td>
<td>0.83</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4.77</td>
<td>6,798</td>
<td>3.34</td>
</tr>
<tr>
<td>SD</td>
<td>0.63</td>
<td>1,669</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*Note. Evaluating time measured in milliseconds (ms).*
mediation effect, we employed the procedure for examining mediation in within-subject design introduced by Judd, Kenny, and McClelland (2001). They indicated that if there is a treatment effect in the outcome variable (e.g., the strength ratings) in within-subject design, a mediator variable (e.g., the evaluating time) can be said to mediate the treatment effect when two conditions are met. First, there must be a significant mean difference between the two treatment conditions in the mediator variable. Second, the mediator variable difference must be predictive of the outcome variable difference. Note that although this procedure is mainly for within-subject variables with two levels, it can be generalized to orthogonal contrasts for within-subject variables with more levels (the argument type in our study has four levels). Thus, we focused on the two orthogonal contrasts which were the contrast between compatible–strong (congruent) and incompatible–strong (incongruent) arguments and the contrast between compatible–weak (incongruent) and incompatible–weak (congruent) ones.

For the first contrast, the previous analyses have revealed that relative to compatible–strong arguments, participants judged incompatible–strong arguments as less extreme and spent more time for them. Next, we, respectively, subtracted the strength ratings and the evaluating time of incompatible–strong arguments from those of compatible–strong arguments to calculate a strength difference and a time difference score. The result further showed that the time difference was a significant predictor of the strength difference, $b < 0.01, t(34) = -4.66, p < .001, \beta = -.63$. Thus, the difference in evaluating time mediated the strength difference between incompatible–strong (incongruent) and compatible–strong (congruent) arguments. For the second contrast, the previous analyses have also revealed that relative to incompatible–weak arguments, participants judged compatible–weak arguments as less extreme and tended to spend more time for them. Next, we also calculated a strength difference and a time difference score between these two arguments. The result showed that the time difference was a marginally significant predictor of the strength difference, $b < 0.01, t(34) = 1.78, p = .085, \beta = .29$. Thus, the difference in evaluating time tended to mediate the strength difference between incompatible–weak (congruent) and compatible–weak (incongruent) arguments. Overall, although the evidence for the mediating role of evaluating time is only marginal for

Figure 2. Interaction of argument compatibility and argument quality on evaluating time. The bars depict one standard error around the mean.
incompatible–weak and compatible–weak arguments, the results generally supported that the processes through which the participants evaluated incongruent arguments as less extreme were related to the extra evaluating time they spent for them.

**Discussion**

The results were consistent with the hypotheses derived from the congruence model. Participants judged incongruent arguments as more neutral in strength and spent more time generating judgments for them compared with congruent arguments. In addition, the mediation analyses also showed that the difference in evaluating time tended to mediate the strength difference between incongruent and congruent arguments. This provided further support for the congruence model that incongruent arguments would cause an extra deliberative reconciliation processing, which may result in them being evaluated as less extreme. However, predictions derived from the disconfirmation model were not fully supported. Although participants did generally rate incompatible arguments as weaker than compatible arguments, they tended to spend more time generating judgments for compatible–weak arguments than incompatible–weak arguments.

**EXPERIMENT 2**

The congruence model indicates that incongruent arguments would elicit intrapsychic conflict causing further reconciliation processing. This point was not examined in Experiment 1. Thus, in addition to evaluating the strength of arguments regarding the announcement of grades of students, participants in Experiment 2 also indicated how hesitant they felt about generating each judgment. First, we predicted that if incongruent arguments elicit intrapsychic conflict and further reconciliations, participants would feel more hesitant generating strength judgments for incongruent (incompatible–strong and compatible–weak) than for congruent (compatible–strong and incompatible–weak) arguments. This prediction runs counter to Edwards and Smith (1996) model. Because they proposed that incompatible arguments that drive individuals to implement refutational analyses involve an extra mental load, participants should feel more hesitant generating judgments for incompatible (incompatible–strong and incompatible–weak) than for compatible (compatible–strong and compatible–weak) arguments. In brief, our model predicts that the subjective hesitation is dependent on both argument quality and argument compatibility, whereas the disconfirmation model predicts that only argument compatibility influences the subjective hesitation.

Next, as in Experiment 1, we predicted that compatible–strong arguments would be judged as the strongest and sequentially followed by incompatible–strong, compatible–weak, and incompatible–weak ones. In addition, Edwards and Smith (1996) found that if the prior beliefs of the participants were accompanied by emotional convictions, which was generally included in the construct of personal importance or attitude importance (e.g., Zuwerink & Devine, 1996), the magnitude and form of the disconfirmation bias was affected (the ratings of incompatible arguments would become weaker, but the ratings of compatible ones would not be different). To extend their findings and provide further evidence for the congruence model, we included personal importance as a predictor variable. Specifically, because the strength judgments for incongruent arguments are made under inner conflict and undergo a reconciliation phase, the nature of the
judgments should be more unstable and more susceptible to change. Moreover, incongruent arguments also require more time and more operating processes which may increase their chance to be influenced. As such, relative to congruent arguments, the strength ratings of incongruent arguments should be more likely to be affected by the perceived personal importance of an issue, which has been shown to enhance a bias to defend one’s attitudes (Bohner & Dickel, 2011; Petty & Cacioppo, 1990) and intensify an individual’s tendency to evaluate arguments depending on argument compatibility (Hart et al., 2009; Petty & Cacioppo, 1979). Thus, we predicted that personal importance would relate to the strength ratings more strongly for incongruent than for congruent arguments. Participants perceiving higher personal importance for the issue would rate compatible–weak (incongruent) arguments as stronger and incompatible–strong (incongruent) ones as weaker. However, the predicted effects of personal importance on the ratings of compatible–strong and incompatible–weak (congruent) arguments should diminish.

Method
Participants and design
A total of 102 undergraduates at a university in western Taiwan received payment and extra course credit for participation. The positions (pro- vs. anti-) and qualities (strong vs. weak) of arguments about the issue of announcement of students’ grades were manipulated, and the participants’ existing positions were measured. The ratings of argument strength and subjective experience of hesitation in generating strength judgments were dependent variables. In addition, we also added one more variable, the personal importance of the issue, as a predictor of the ratings of argument strength.

Materials
‘Should teachers announce every student’s grade to the class after an examination’ was an issue prominent in Taiwan at the time the study was run. We first drafted 14 arguments supporting and 16 arguments opposing the position. Four types of arguments were generated, such as ‘the announcement of grades can stimulate students to study hard’ (a pro-strong argument), ‘the announcement of grades can promote the relationship between students and the school’ (a pro-weak argument), ‘the concealment of grades can lead students to not value scores very much’ (an anti-strong argument), and ‘the concealment of grades can enhance students’ identification with school’ (an anti-weak argument).

We handed these arguments over to the same five experts and chose 16 final arguments. The median correlation between the strength ratings of the six experts (including one of the authors) was .69. The median correlation between the five external experts was .63 ($p < .05$). The mean strength ratings for pro-strong, pro-weak, anti-strong, and anti-weak arguments were 5.04, 2.54, 5.21, and 2.63, respectively. A repeated-measures ANOVA revealed a main effect of argument type, $F(3, 15) = 22.42, p < .001, \eta^2 = .82$. There were no significant differences between pro-strong and anti-strong arguments ($p = .465$) or between pro-weak and anti-weak arguments ($p = .773$). However, pro-strong and anti-strong arguments were both rated as stronger than pro-weak ($ps = .006, .008$) and anti-weak ones ($ps = .003, .002$).
In addition, 78 undergraduates from two universities in northern Taiwan rated the strength for each argument. A two-way ANOVA showed no main effect for argument position, $F(1, 77) = 0.33, p = .569$, nor did argument position interact with argument quality, $F(1, 77) = 0.08, p = .779$. However, strong arguments were recognized as stronger ($M = 4.20, SD = 0.65$) than weak ones ($M = 2.59, SD = 0.67$), $F(1, 77) = 329.51, p < .001, \eta^2 = .81$.

Procedure and measuring instruments
The procedure was similar to Experiment 1 with the following exceptions: First, evaluating time was not recorded. All the materials employed were presented on successive pages of a packet. Second, participants rated the strength for each of the 16 arguments arranged in one of three random sequences while at the same time indicating how hesitant they felt about generating each judgment ($1 = none at all, 6 = a great deal$). Third, in addition to some filler items and one question identical to that in Experiment 1 (i.e., the precise position), participants answered two more questions regarding the perceived personal importance of the issue: That is, ‘How important is this issue to you personally?’ ($1 = not at all important, 6 = very important$) and ‘How emotionally concerned are you about this issue?’ ($1 = none at all, 6 = a great deal$). Individuals’ responses were averaged to form a personal importance index with acceptable reliability ($\alpha = .68$).

Results
Preliminary analyses
Both the pre-judgment position ($M = 1.56, SD = 0.50$) and post-judgment position ($M = 3.62, SD = 1.34$) of the 102 participants were slightly favourable towards pro-announcement. The correlation between the pre- and post-judgment positions was .83 ($p < .001, N = 102$). There were five participants who did not keep the same position in the pre- and post-judgment stage. After excluding their data, the correlation between the pre- and post-judgment positions was .88 ($p < .001, N = 97$).

Tests of predictions
After excluding a further participants’ data because of incomplete answers on the measurement of the subjective hesitation, a two-way ANOVA revealed that there was no main effect for argument compatibility on the subjective hesitation, $F(1, 95) = 0.36, p = .549$. However, the main effect of argument quality and their interaction were both significant, $F$s (1, 95) $> 12.47, \eta^2 > .12$ (see Figure 3). Overall, participants felt more hesitant generating judgments for weak ($M = 2.47, SD = 0.65$) than strong arguments ($M = 2.20, SD = 0.50$). However, simple effects analysis revealed that participants were more hesitant generating judgments for compatible–weak ($M = 2.54, SD = 0.78$) than incompatible–weak arguments ($M = 2.41, SD = 0.66$), $F(1, 190) = 4.17, p = .043, \eta^2 = .02$. They were also more hesitant generating judgments for incompatible–strong ($M = 2.29, SD = 0.59$) than compatible–strong arguments ($M = 2.10, SD = 0.58$), $F(1, 190) = 8.41, p = .004, \eta^2 = .04$.

Next, we found a main effect of argument type on the strength ratings, $F(3, 288) = 117.74, p < .001, \eta^2 = .55$. As in Experiment 1, compatible–strong arguments
were rated as strongest ($M = 4.72$, $SD = 0.69$) and sequentially followed by incompatible–strong ($M = 4.11$, $SD = 0.84$), compatible–weak ($M = 3.45$, $SD = 0.84$), and incompatible–weak ones ($M = 2.99$, $SD = 0.86$). In addition, we predicted that personal importance would relate to the strength ratings more strongly for incongruent arguments than for congruent arguments. To examine this moderation effect, we employed the procedure for examining moderation in within-subject design introduced by Judd et al. (2001). They indicated that to assess whether the magnitude of a treatment effect in the outcome variable (e.g., the strength ratings) is moderated by a stable concomitant variable (e.g., the personal importance) in within-subject design, the outcome variable difference should be regressed on that concomitant variable. If the magnitude of the treatment effect depends on the concomitant variable, then the concomitant variable will be predictive of the outcome variable difference. As in Experiment 1, we focused on the same two orthogonal contrasts to test this moderation effect.

For the first contrast between compatible–strong (congruent) and incompatible–strong (incongruent) arguments, the strength difference was regressed on the personal importance. The result showed that the slope for personal importance was significantly different from zero, $b = 0.15$, $t(95) = 3.10$, $p = .003$, $\beta = .30$. This showed that the slope for personal importance in predicting the strength ratings of compatible–strong arguments differed from the slope in predicting incompatible–strong arguments (see Figure 4). A simple slopes analysis showed that the main effect of personal importance on the ratings of incompatible–strong arguments was significant, $b = -0.10$, $t(95) = -2.20$, $p = .030$, $\beta = -.22$. Participants who perceived a higher personal importance for the issue rated incompatible–strong arguments as weaker. However, the main effect of personal importance on the ratings of compatible–strong arguments was not significant, $b = 0.05$, $t(95) = 1.44$, $p = .150$, $\beta = .15$. For the second contrast between compatible–weak (incongruent) and incompatible–weak (congruent) arguments, the slope for personal importance in predicting the strength difference was also significantly different from zero, $b = 0.11$, $t(95) = 2.12$, $p = .037$, $\beta = .21$ (see also Figure 4). A simple slopes analysis showed that the main effect of personal importance

![Figure 3. Interaction of argument compatibility and argument quality on subjective hesitation. The bars depict one standard error around the mean.](image-url)
on the ratings of compatible–weak arguments was significant, $b = 0.09$, $t(95) = 2.05$, $p = .044$, $\beta = .21$. Participants who perceived a higher personal importance rated compatible–weak arguments as stronger. However, the main effect of personal importance on the ratings of incompatible–weak arguments was not significant, $b = -0.02$, $t(95) = -0.29$, $p = .780$, $\beta = -.03$. Overall, the results showed that the predicted effects of personal importance on the strength ratings tended to be stronger for incongruent than for congruent arguments.

### Additional analysis

The results showed that participants tended to feel more hesitant generating judgments for incongruent arguments and judged them as less extreme in strength than congruent ones. However, if participants evaluated incongruent arguments as less extreme primarily because incongruent arguments elicited intrapsychic conflict and further reconciliations, the subjective hesitation should mediate the effect of argument congruency on the strength ratings. To test this mediation effect, we focused on the same two orthogonal contrasts. For the first contrast between compatible–strong and incompatible–strong arguments, the result showed that the hesitation difference was a significant predictor of the strength difference, $b = -0.54$, $t(94) = -3.62$, $p < .001$, $\beta = -.35$. For the second contrast between compatible–weak and incompatible–weak arguments, the hesitation difference was a marginally significant predictor of the strength difference, $b = 0.24$, $t(94) = 1.64$, $p = .105$, $\beta = .17$. Overall, the results showed that the difference in subjective hesitation tended to mediate the strength difference between incongruent (incompatible–strong and compatible–weak) and congruent (compatible–strong and incompatible–weak) arguments.

Finally, it is noteworthy that personal importance was a moderator affecting the magnitude of the effect of argument congruency on the strength ratings. However, the subjective hesitation could be seen as a mediator that mediated the effect of argument congruency on the strength ratings. According to Edwards and Lambert (2007), it is not necessary that personal importance (a moderator) must have a main effect or an interactive effect (e.g., personal importance × argument congruency) on the subjective
hesitation (a mediator) when we argued that personal importance moderated the effect of argument congruency on the strength ratings. Thus, we did not make a prediction about this. However, because this issue may be of interest to readers, we also examined the effect of personal importance on the subjective hesitation. The results showed that personal importance did not reveal any main or interactive effects on the subjective hesitation. The main effects of personal importance on the participants’ subjective hesitation in evaluating compatible–strong, incompatible–strong, compatible–weak, and incompatible–weak arguments were all non-significant, $bs < 0.10$, $ts (94) < 1.09$, $ps < .281$, $bs < .11$.

**Discussion**

The results were consistent with our predictions. Participants tended to feel more hesitant generating judgments for incongruent arguments and judged them as more neutral in strength than congruent ones. In addition, the difference in the subjective hesitation tended to mediate the strength difference between incongruent and congruent arguments. These results support the conclusion that incongruent arguments should elicit intrapsychic conflict and undergo an extra reconciliation processing, which may result in incongruent arguments being evaluated as less extreme. However, predictions made from the disconfirmation model were not fully supported. Although participants did generally rate incompatible arguments as weaker than compatible ones, they felt more hesitant generating judgments for compatible–weak than for incompatible–weak arguments.

In addition, the results also indicated that the effects of personal importance on the strength ratings of compatible–strong and incompatible–weak (congruent) arguments were not significant. However, participants perceiving a higher personal importance for the issue tended to judge compatible–weak (incongruent) arguments as stronger and incompatible–strong (incongruent) ones as weaker. This indicates that people with higher personal importance for the issue more easily see the merits of compatible–weak arguments and the demerits of incompatible–strong arguments (instead of the merits or demerits of the other two congruent arguments). These findings further support the notion that because incongruent arguments undergo a reconciliation phase, the strength ratings for these arguments would be more susceptible to change. These arguments also require more time and more operating processes which may increase their chance to be influenced. Thus, their strength ratings would be more likely affected by the personal importance.

**GENERAL DISCUSSION**

Results in the two experiments show that incongruent arguments are judged to be less extreme in strength (in both experiments), are scrutinized longer (in Experiment 1), and cause participants to feel more hesitant generating strength judgments for them (in Experiment 2). In addition, the strength ratings of incongruent arguments are more closely associated with the perceived personal importance of the issue (in Experiment 2). These findings indicate that incongruent arguments (not compatible arguments) should undergo an extensive reconciliation processing (not a refutational processing) and thus support the argument evaluating processes delineated by the congruence model (instead of the disconfirmation model). Incongruity between evaluating tendencies from argument compatibility and argument quality, and not simply the argument compatibility,
plays a more important role in activating an extensive processing in the evaluation of arguments.

Interestingly, the findings of Edwards and Smith (1996), including those supportive or not supportive of the disconfirmation model, seem to be explained well by our study. Specifically, they predicted that incompatible arguments take longer to be evaluated, are rated as weaker, and cause the participants to generate more thoughts and more refutational (rather than supportive) materials than compatible ones. They tested these predictions in a study that used 14 arguments of seven issues. Confirmation was fully obtained on four of these issues. They also found that when the participants’ prior beliefs were accompanied by emotional convictions, the ratings of incompatible arguments became weaker, but the ratings of compatible ones were not different.

Among the four issues which mostly support the disconfirmation model (death penalty, strike child, parental consent/abortion, and gay/lesbian adoptions), after collapsing the pro- and anti-arguments (see Edwards & Smith, 1996, p. 10), the averaged argument quality on the 7-point scale were, respectively, 3.85, 3.65, 3.72, and 3.65 – all above the averaged rating of 14 arguments used in their study ($M = 3.55$). However, among the other three issues mostly inconsistent with their predictions, the averaged ratings were, respectively, 3.38, 3.43, and 3.20 – all below the averaged rating. That is, among the four issues mostly supportive of the disconfirmation model, the arguments used seem more like compatible–strong and incompatible–strong arguments. Based on the congruence model, because the latter were incongruent arguments, evaluation should take longer than for the former, as well as be rated as weaker, cause more thoughts, and be more easily influenced by emotional conviction. On the contrary, among the three issues mostly not supportive of the disconfirmation model, the arguments used seem more like compatible–weak (incongruent) and incompatible–weak (congruent) arguments. Thus, it would not be necessary for the incompatible (but actually congruent) arguments to take more time, to be rated as weaker, or to cause more thoughts than compatible (but actually incongruent) arguments. In summary, it is possible that incompatible arguments would cause extensive processing only when they are strong enough arguments to elicit contradicting evaluating tendencies. If they are not strong enough, there will be no extensive processing because there is no inner conflict to reconcile.

**Relationship to other works**

Some studies on selective exposure have indicated that people tend to selectively avoid incongruent (compatible–weak and incompatible–strong) arguments and expose themselves to congruent (compatible–strong and incompatible–weak) ones (Hart *et al.*, 2009; Kleinhesselelink & Edwards, 1975). This provides evidence that incongruent arguments do cause intrapsychic conflict, which leads people to ignore these arguments if they can. In the current study, however, if these arguments are evaluated and cannot be avoided, we argue that such inner conflict leads to more extensive processing. This is consistent with research, indicating that various sorts of incongruities in a message can lead to enhanced information processing (e.g., Baker & Petty, 1994; Clark *et al.*, 2012; Maheswaran & Chaiken, 1991).

Previous studies suggested that individuals can simultaneously hold both positive and negative evaluations towards the same object and that such attitudinal ambivalence causes an individual to subjectively feel uncomfortable (Bassili, 2008; Conner & Armitage, 2008), motivates the individual to resolve intrapsychic conflict (Festinger, 1957; Newby-
Clark et al., 2002), and further influences the processing of the received information and the persuasiveness of the argument (e.g., Clark et al., 2008; van Harreveld, van der Pligt, de Vries, Wenneker, & Verhue, 2004). In general, the current study has a similar rationale to these studies, emphasizing that the individual has to implement extra processing to resolve inner conflict elicited by contradicting tendencies. There are, however, several differences. First, studies of attitudinal ambivalence are more concerned with the persuasion effect or the processes of attitude change related to attitudinal ambivalence. This study is concerned mainly with the processes of argument evaluation. Second, they focus primarily on the conflict of contradicting evaluations which the individual simultaneously holds towards the same issue before he receives subsequent information to process. This study, however, focuses mainly on the conflict resulting from contradicting tendencies which are immediately elicited by different attributes of an argument while the individual is evaluating it. Finally, when holding ambivalent attitudes, the feeling of discomfort is assumed as the key potency to give rise to further processing. However, this study indicates that the inner conflict from contradicting evaluating tendencies activates further reconciliations and does not take an explicit position on the activating potency. Future research is needed to explore constructs such as whether a feeling of discomfort or cognitive uncertainty are key potencies that cause further reconciliation processing when evaluating incongruent arguments.

In addition, although several well-known cognitive response models, such as the elaboration likelihood model (see Petty & Wegener, 1999) or the heuristic–systematic model (see Chen & Chaiken, 1999), are mainly concerned with attitude change processes, whereas the current study is concerned with argument evaluation, there are several points worth comparing. First, similar to our notion that people can simultaneously be rational and irrational (biased) while evaluating argument strength, these cognitive response models also make a distinction between relatively objective processing versus biased processing and argue that both processing co-occur while generating an attitude. Next, the cognitive response models generally agree that if people are affected by other factors that increase their motivation or ability to process the persuasive message more deeply, the effect of the argument quality on persuasion will be increased (e.g., Clark et al., 2008; Petty & Wegener, 1999). However, according to the congruence model, the amount of processing that people engage in to evaluate arguments does not necessarily increase the effect of argument quality on strength ratings. When people are driven to extensively process the incongruent arguments due to intrapsychic conflict, strength ratings would indicate a more reconciliatory conclusion, instead of being made more dependent on the objective quality of the argument.

**Practical implications, limitations, and future directions**

The results indicate some noteworthy points that can help people be more rational when sharing opinions with others. Our results showed that when we are evaluating the strength of arguments, although rational judgments can be based on the objective quality of the arguments, judgments may also be biased by prior beliefs. Previous studies have indicated that if people are instructed to correct the tendency to process information depending on the prior beliefs, it is possible to reduce this bias (e.g., Kunda, 1990; Liu, Chen, Chang, & Chang, 2011; Stanovich & West, 1997). Thus, to anyone who is going to evaluate arguments, it is important to remind oneself to monitor the influence of prior beliefs on the judgments. In addition, if the issue is perceived to be personally important, it is better to remind oneself that the biased
evaluating tendency can be augmented, especially when the arguments are compatible–weak or incompatible–strong arguments.

The current study has several limitations. First, we invited experts to rate each argument’s strength to assure differentiation of arguments’ strength. Although the median correlations between the strength ratings of six experts were both significant in two experiments, they were not high (especially in Experiment 1, \( r = .58, p < .05 \)). In addition, there were only a limited number of arguments used in two experiments (only four arguments for each type) and the sample size was not large in Experiment 1 (\( n = 40 \)). Thus, to replicate and enhance comprehensiveness and generalizability of the results, future studies further assuring the differentiation of the arguments’ strength and including more experiments, arguments, and participants are needed.

The current study is also related to several issues worthy of future investigation. First, although several early studies of the effect of attitude congeniality on memory suggested that memory is better for congenial (compatible) than for uncongenial (incompatible) information (e.g., Levine & Murphy, 1943), many later studies have failed to replicate the congeniality effect (e.g., Greenwald & Sakumura, 1967) and even occasionally reported opposite results (e.g., Cacioppo & Petty, 1979). Eagly et al. (2001) suggested that although compatible information could be memorized well because of its superior fit with existing attitudes, its inherent pleasantness, and other advantages it may possess, the congeniality effect is difficult to detect because people actively counterargue and thus enhance memory for incompatible information. If one of the key factors determining the congeniality effect is whether extensive processing of information is implemented, our perspective regarding the argument evaluation processes may be helpful in understanding this long-standing issue.

Next, although the congruence model focuses on the congruence between evaluating tendencies elicited by argument compatibility and argument quality, other factors that have been shown to influence persuasion effects also seem worth considering. These factors may elicit evaluating tendencies in the evaluation of arguments, such as source expertise (Clark et al., 2012) and source likability (Chaiken, 1980). Whether the mechanisms formulated by the congruence model apply to such a complicated situation would be an interesting question.

Finally, when someone is going to make decisions or to execute actions in everyday life, it is possible they also hold different reacting tendencies at the same time. Thus, the congruence among these tendencies may also play an important role in the underlying processes of the individual’s decisions or behaviours. For example, in a study regarding guilt judgment, Alison et al. (2012) pointed out that if the profile of the possible offender made by the participant contradicts the unorthodox profile made by the expert (relative to the orthodox profile), the participant is more likely to re-evaluate the guilt judgments of the suspects. According to the congruence model, this implies that the incongruence between the participant’s prior belief and the expert advice may cause a more extensive processing. In sum, the congruence model need not be confined to the scope of argument evaluation and may feasibly be generalized to explain people’s judgments and behaviours in other fields.

Acknowledgement

This research is supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant No. NSC 99-2410-H-364-003 and Grant No. NSC 100-2410-H-364-007- MY2.
References


Received 8 February 2013; revised version received 14 January 2015