Attenuating Belief Bias Effects in Syllogistic Reasoning:
The Role of Belief-Content Conflict

by

Michelle Colleen Elizabeth Hilscher

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Psychology
University of Toronto
© Copyright by Michelle Colleen Elizabeth Hilscher 2014
Attenuating Belief Bias Effects in Syllogistic Reasoning: The Role of Belief-Content Conflict

Michelle Colleen Elizabeth Hilscher

Doctor of Philosophy

Psychology
University of Toronto

2014

Abstract

A reasoner’s beliefs can compromise or inflate the accuracy of their syllogistic judgments when syllogistic content and structure are incongruent or congruent, respectively. An integrated approach to the study of syllogistic reasoning led to the investigation of premise-based belief-content conflict and its impact on belief bias. The belief-content conflict cue attenuated belief bias in incongruent valid and invalid trials, as well as congruent invalid trials. Its efficacy was found to depend on the difficulty level of the syllogism in which it was embedded and the location of its placement. Reaction time analyses were used to guide interpretations about the relative engagement of Systems 1 and 2. The key findings suggested that belief-content conflict activated System 2 for invalid incongruent trials which would otherwise have been processed using low-cost heuristic means due to their inherent difficulty. In contrast, it appeared that in valid trials the cue led to a redirection of System 2 resources such that specialized analytic strategies were applied in incongruent trials preceded by belief-content conflict compared to those lacking this cue. Finally, belief bias was successfully offset by belief-content conflict even in cases of congruency. In congruent invalid trials without this cue participants’ intuitive awareness of the content-structure match appeared to lead to low-cost, belief-based guesses; yet when presented as the major premise this conflict cue appeared to shift System 1 processing away from content and towards structure. Albeit less diligent than System 2 analysis, the shallow
consideration of structural features may have been viewed as a safer bet than any shortcut aiming to capitalize on syllogistic content. This set of findings cannot be fully accounted for by the selective scrutiny, misinterpreted necessity, mental models, verbal reasoning, selective processing, or Receiver Operating Characteristics accounts thereby highlighting the need for considering belief-content conflict in future models of belief bias.
Acknowledgements

I would like to express my gratitude to Gerry for his support and encouragement throughout the years we have worked together. Above all else, you have inspired in me an enduring interest in the history and philosophy of science. Thank you Oshin and Morris for your guidance; together you have ensured that I know the little details and also see the big picture. Finally, to Arthur, my family, and my friends – thank you for being wonderful and never losing faith in my ability to complete this process!
# Table of Contents

## Chapter 1: The Syllogistic Reasoning Paradigm
1

- The logical syllogism
  - Structural variables
  - Content variables
- The typical logical syllogism task
- Why do syllogistic reasoning errors occur?
  - The atmosphere effect
  - Ambiguity and caution
  - The conversion error
  - Probabilistic reasoning
  - The belief bias effect
- Syllogistic reasoning errors may not reflect flawed reasoning
  - The figural effect
- Elaborating the role of structural variables
  - Which is dominant? Atmosphere or conversion?
  - What about figure, does it factor in as a source of syllogistic reasoning errors?
- Summary: Structural variables

## Chapter 2: Reassessing the Relationship Between Structure and Content
21

- The historical polarity of structure and content
- New Criticism
- History of psychology
- An attempted rapprochement of structure and content
- Elaborating the role of content variables
  - Establishing that belief bias is a genuine phenomenon
  - Conditions that inspire belief bias
    1. emotional content
    2. emotional frame of mind
    3. emotional content and emotional reasoners
    4. working memory and executive functioning
    5. other personal qualities of the reasoner
    6. argument variables
    7. source credibility
- Why belief bias happens
- Reaching a consensus by the eighties?

## Chapter 3: Syllogistic Reasoning Models
53

- The selective scrutiny model
- The misinterpreted necessity model
- The mental models account
- Verbal reasoning theory
- Selective processing account
- Receiver operating characteristics account
- Comparing the models
<table>
<thead>
<tr>
<th>Chapter 4: Developing an Integrated Approach to Belief Bias</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconciling the models</td>
<td></td>
</tr>
<tr>
<td>Multiple analytic processes are possible</td>
<td></td>
</tr>
<tr>
<td>Analytic processing is distinct from heuristic processing</td>
<td></td>
</tr>
<tr>
<td><em>The heuristics system</em></td>
<td></td>
</tr>
<tr>
<td>Evidence for dual-process systems</td>
<td></td>
</tr>
<tr>
<td><em>Heuristic and analytic system interaction</em></td>
<td></td>
</tr>
<tr>
<td>i. system 1 predominates</td>
<td></td>
</tr>
<tr>
<td>ii. systems 1 and 2 are processing partners</td>
<td></td>
</tr>
<tr>
<td>iii. system 2 can act on system 1</td>
<td></td>
</tr>
<tr>
<td>iv. system 1 and 2 process the same information, albeit in discrete ways</td>
<td></td>
</tr>
<tr>
<td>v. system 2 specialists are a distinct ‘type’ of person</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5: The Current Project</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content, context, and beliefs can improve analytical strategies</td>
<td></td>
</tr>
<tr>
<td>Premise plausibility</td>
<td></td>
</tr>
<tr>
<td>Purpose of current studies</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6: Belief-Content Conflict Attenuates Belief Bias</th>
<th>115</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1A</td>
<td></td>
</tr>
<tr>
<td><em>Rationale and expectations</em></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td><em>Does the ‘warning signal’ effect translate into ensuing trials?</em></td>
<td></td>
</tr>
<tr>
<td><em>Summary of findings and preliminary interpretations</em></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7: Belief-Content Conflict Impacts Semantic Processing</th>
<th>142</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1B</td>
<td></td>
</tr>
<tr>
<td><em>Rationale and expectations</em></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td><em>Summary of findings and preliminary interpretations</em></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 8: Efficacy of Belief-Content Cue Depends on Syllogistic Difficulty Level</th>
<th>157</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 2</td>
<td></td>
</tr>
<tr>
<td><em>Rationale and expectations</em></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 9: Belief-Content Conflict Can Inspire Conclusion-Content Blocking</th>
<th>179</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 3</td>
<td></td>
</tr>
<tr>
<td><em>Rationale and expectations</em></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td><em>Evaluating semantic plausibility – Summary of findings and preliminary interpretations</em></td>
<td></td>
</tr>
</tbody>
</table>
## Chapter 10: Discussion

<table>
<thead>
<tr>
<th>Belief-content conflict attenuates belief bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid trials</td>
</tr>
<tr>
<td>Invalid trials</td>
</tr>
<tr>
<td>System 1 versus System 2</td>
</tr>
<tr>
<td>i. strategy shifting and activation</td>
</tr>
<tr>
<td>ii. inhibition</td>
</tr>
<tr>
<td>The six models of reasoning</td>
</tr>
<tr>
<td>Extension beyond the six models</td>
</tr>
</tbody>
</table>

Conclusion

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Appendices</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>End Note</th>
</tr>
</thead>
</table>
CHAPTER 1:  
THE SYLLOGISTIC REASONING PARADIGM

Reasoning and the evaluation of logical syllogisms have long been celebrated activities. In the 6th Century BC, the Formalists of Asia Minor attempted to deduce the ultimate nature of the universe by reasoning about mathematical rules. Plato instructed his pupils to deduce truths about life from premises that were assumed to be innate and universally accessible. Following in this vein, the German Rationalists of whom Gottfried Leibniz (1646-1716) was one, elaborated on the nature of this innate knowledge, proposing that number, space, time, substance, and causality are inherent contents of the mind. George Berkeley (1685-1753) went so far as to claim that the world is created in an entirely top-down fashion; thus deduction and perception are one and the same (Benjafield, 2010).

Historically speaking, logical thinking and syllogistic reasoning were often viewed as morally superior to emotional experience and intuition. It was assumed by numerous scholars that reasoning was the process that would yield the *truest product* – sound knowledge about life and the universe. Despite this philosophical orientation towards the *product* of reasoning, experimental research on syllogistic reasoning has mostly been aimed at uncovering how reasoning occurs – an orientation towards the *process*. For instance, researchers in the 1930s through 1960s debated the conditions under which accurate and inaccurate syllogistic reasoning are likely (e.g., Woodworth & Sells, 1935; Janis & Frick, 1943; Chapman & Chapman, 1959). Janis and Frick (1943) proposed that reasoning may go off track when reasoners are swayed by their beliefs; this was the first time that the belief bias effect was proposed. Janis and Frick’s (1943) work was followed up by two streams of research. The first stream assumed that belief bias was a bonafide phenomenon and attempted to elaborate on the conditions that might make it
likely to occur (e.g., Morgan & Morton, 1944; Lefford, 1946). The second stream studied sources of reasoning error other than belief bias and implicitly challenged the existence of this phenomenon (e.g., Simpson & Johnson, 1966; Begg & Denny, 1969). In 1983, Evans, Barston and Pollard weighed in on this debate by publishing an article in which they concluded that the belief bias effect exists and that it is indeed a valid explanation for errors in syllogistic reasoning. More recent syllogistic reasoning research marks a splintering in the field as scholars have proposed competing models of reasoning that offer varying explanations for syllogistic reasoning generally, and belief bias more specifically (e.g., selective scrutiny; misinterpreted necessity; mental models account; verbal reasoning theory; selective processing theory; Receiver Operating Characteristics account). Current disagreements about how beliefs are thought to impact syllogistic reasoning may be attributed to the fact that there is currently a lack of consensus about the latter process (Khemlani & Johnson-Laird, 2012).

Given this historical framework I would like to make the scope of my own work clear. I have designed a set of logical syllogisms that have allowed me to replicate the belief bias effect. Besides allowing for a replication of this classic phenomenon, my studies highlight the role of a novel factor, premise-based belief-content conflict, in syllogistic reasoning. This factor is shown to mediate the belief bias effect stemming from belief-logic conflict in the conclusion. Yet, the current models that account for syllogistic reasoning cannot explain why an implausible premise attenuates belief bias. Thus, a key goal of my dissertation is to develop an integrated perspective about reasoning and the belief bias effect which adequately accounts for the role of belief-content conflict and premise-based reasoning.

The purpose of this introduction is to establish a foundation for my dissertation work by describing the structural and content variables associated with the logical syllogism, explaining
the typical syllogistic task, and highlighting known antecedents of belief bias. These sections will be developed in keeping with a critically and historically grounded philosophy of science. Consequently, the ways that classical scholars and contemporary researchers have defined and studied reasoning will be contextualized in a sociocultural manner. Against this backdrop, I will summarize and attempt to reconcile six competing models of syllogistic reasoning that are generally held to be exclusive accounts. I will highlight commonalities between these models, namely that there is evidence that both verbal and spatial reasoning occur, and that the analytic versus heuristic distinction is a foundation of each approach. Finally, I will pose some main questions that represent common shortcomings in these models pertaining to the role of context and beliefs generally, as well as plausibility and premise-based semantics more specifically. These questions are the inspiration for my own work and are at the core of the four studies described in this dissertation.

**The Logical Syllogism**

*Structural Variables*

The logical syllogism is a three part deductive argument that consists of a major premise, minor premise, and conclusion. The major premise is the first statement of the argument and it makes a categorical proposition about the predicate in the conclusion. The minor premise is the second statement and it is a categorical proposition that introduces the subject of the conclusion. Each proposition has a *quantity* (universal or particular) and a *quality* (affirmative or negative), thus there are four possible propositions: **A** - universal and affirmative (e.g., all X are Y); **E** - universal and negative (no X are Y); **I** - particular and affirmative (some X are Y); or **O** - particular and negative (some X are not Y). Combining three propositions forms an argument that possesses *validity*. Specifically, the structure of a logical syllogism is valid if the major and
minor premises imply that the conclusion is “true”. Alternately, the logical syllogism is invalid if the conclusion is not a necessary product of the major and minor premises.

*Content Variables*

Propositions within a syllogistic argument can vary in terms of their *meaning*, *emotionality*, and *plausibility*. In terms of its meaning, a proposition can be non-semantic (i.e., with abstract terms like X and Y) or semantic (i.e., with concrete terms like cat and dog). If the proposition is semantic then it can be further described as either neutral (i.e., facts-based and unemotional, for instance, all dogs are mammals) or emotional (i.e., tapping into beliefs and feelings, for instance, all pit bulls are dangerous). Finally, a proposition is plausible if it presents information that is consistent with real-world knowledge (e.g., some dogs are brown) or that represents a possible reality (e.g., some planets are earth-like). Alternately, a proposition is deemed implausible when it makes a claim that is empirically unsound (e.g., all dogs are brown; all planets are earth-like).

**The Typical Logical Syllogism Task**

Researchers have typically studied syllogistic reasoning by presenting a three part argument and asking reasoners to determine whether the premises necessarily imply the conclusion or by presenting two premises and asking reasoners to formulate a valid conclusion. Regardless of the method employed, it is not surprising that reasoners often misjudge the validity of the syllogism, thereby committing the hallmark error that has led to much consideration and debate by the syllogistic reasoning research community.

**Why do Syllogistic Reasoning Errors Occur?**

Different explanations have been offered to account for the inaccurate judgment of syllogistic validity. In the coming section the core hypotheses that spurred debate and
independent streams of research in the field will be compared. Woodworth and Sells (1935) proposed that atmosphere (1) as well as ambiguity and caution (2) lead to syllogistic reasoning errors. Chapman and Chapman (1959) countered with their conversion (3) and probabilistic reasoning (4) hypotheses. Janis and Frick (1943) questioned the role of semantics and belief bias (5). Henle (1962) presented the somewhat radical view that syllogistic reasoning may fall prey to pre-logical errors or, a failure on the part of the individual to commit to the reasoning process altogether (6). Finally, Frase (1968) initiated a discussion about the role of syllogistic figure which is determined by the position of the predicate and subject within the major and minor premises. Frase proposed that certain figures are more inaccurately evaluated than others (7).

Some of these explanations exclusively implicate structural variables (e.g., the atmosphere effect is a product of certain premise types and it is the source of syllogistic reasoning errors), whereas others blame poor syllogistic reasoning on the content of the argument (e.g., the belief bias effect occurs when the conclusion makes an assertion about the real-world). Upon reflection, it seems as though many explanations emphasize structure over content, or vice versa. It is not my intention to continue along this either-or vein of theorizing because there is evidence that would suggest that both structural and content variables play a role. Thus, my approach will address the relationship between structure and content – how do these variables interact and impact syllogistic reasoning?

The Atmosphere Effect

Woodworth and Sells (1935) proposed that the atmosphere of a syllogism can encourage inaccurate reasoning. They hypothesized that an atmosphere is created by the quantity (universal or particular) and quality (affirmative or negative) of major and minor premises and that
participants are more likely to accept a conclusion that is in kind with the atmosphere inspired by these preceding propositions. The general rules that arise from this hypothesis are:

1. If one premise is particular (some are; some are not) this encourages the reasoner to accept a particular conclusion.
2. If one premise is negative (none are; some are not) this encourages the reasoner to accept a negative conclusion.

Sells (1936) went on to document patterns of error that support the atmosphere hypothesis; specifically that false positives (asserting that an invalid conclusion is valid) arose in a pattern that was predicted by the atmosphere of the premises (e.g., an invalid argument that ends with a particular negative conclusion will be misjudged to be valid when the premises are particular and negative) (as cited in Chapman and Chapman, 1959).

Ambiguity and Caution

Woodworth and Sells (1935) supposed that sometimes syllogistic errors arise, not because of an atmosphere effect, but rather because of the ambiguity of the word “some,” and caution on the part of the reasoner. Left to their own devices to interpret the qualifier “some,” a reasoner may decide (albeit erroneously) that the assertion “some x are y” implies that “some x are not y’. A reasoner may also respond in a naturally conservative manner which leads them to err on the side of caution by accepting particular (as opposed to universal) and negative (as opposed to affirmative) propositions. As Woodworth and Sells (1935) put it, participants’ wariness favors “the acceptance of weak and guarded rather than strong conclusions” (p. 460).

The Conversion Error

Chapman and Chapman (1959) tested the atmosphere hypothesis by administering invalid syllogisms to participants who had to select the appropriate conclusion from a number of
possible alternatives. The correct response for every trial happened to be “none of the above are the correct conclusion,” and the accuracy in this task was very low (only 20% of the trials were accurately judged). Despite the low level of accuracy and consequent abundance of errors to analyze, the pattern of errors did not fit Woodworth and Sells’ hypothesis. Consequently, Chapman and Chapman suggested that the pattern of errors documented by Woodworth and Sells was a product of methodology rather than the consequence of an atmosphere effect. Woodworth and Sells (1935) had employed a binary scoring system where “absolutely true” and “probably true” were scored as agreement, and “indeterminate” and “absolutely false” were treated as instances of disagreement. According to Chapman and Chapman, this scoring system artificially ensured that I (Some X are Y) and O (Some X are not Y) responses would be more likely than A (All X are Y) and E (No X are Y) responses. Woodworth and Sells believed that reasoners tended to err on the side of I and O because these conclusions mirrored the atmosphere of the premises. Yet, according to Chapman and Chapman, reasoners were in reality no more likely to pick I and O over A and E, it was the scoring system that created this illusion.

In the absence of corroborating evidence for the atmosphere hypothesis, Chapman and Chapman (1959) proposed that the pattern of errors, might be explained by way of two hypotheses – firstly, that participants commit the error of conversion (assuming that Some X are Y implies that Some Y are X), and secondly, that participants engage in probabilistic reasoning. In line with the former, the authors suggested that the conversion error may stem from habits that reasoners have derived from the study of mathematics where they have learned to equate variables as opposed to classifying them. In the case of a syllogistic argument, Equating X with Y (X=Y) when solving a syllogism is a mathematically derived strategy that breaks logical rules which require that X be seen as belonging to some degree to the Y category.
**Probabilistic Reasoning**

The second explanation offered by Chapman and Chapman was that participants use *probabilistic reasoning* in which they mistakenly conclude that because two terms (X and Y) have a quality in common (Z), it is relatively likely that they occupy a common category themselves.

**The Belief Bias Effect**

Despite controlling for the purported effects of atmosphere, Janis and Frick (1943) reported that reasoners continued to commit logical errors when attempting to solve syllogisms. The syllogisms being studied by Janis and Frick were *semantic* and they proposed that it was the conflict between content and structure that was at the root of syllogistic reasoning errors. When a conclusion was deemed to be empirically implausible (e.g., the copperhead is not a poisonous snake), participants were more apt to conclude that the syllogism is invalid, regardless of its actual structure.

**Syllogistic Reasoning Errors May Not Reflect Flawed Reasoning**

Theorists such as Woodworth and Sells (1935), Chapman and Chapman (1959) and Janis and Frick (1943) supposed that certain variables disrupt logical processing such that syllogistic reasoning errors represent *failures in logic*. Contrary to these theorists, Henle (1962) proposed that syllogistic reasoning errors occur because participants commit errors *before reasoning begins*, or fail to commit themselves to *reasoning logically* in the first place. As such, errors on a reasoning task may not necessarily reflect errors *in reasoning*. Based on data obtained through the administration of a syllogistic task as well as post-task interviews, Henle accounted for inaccurate syllogistic evaluations in the following ways:
a. The task may not have been understood; therefore syllogistic thinking may not have been attempted. In particular, participants may have focused mistakenly on plausibility instead of validity.

b. The ingredients may have been changed by the participant. If one of the premises or the conclusion was unwittingly adjusted (e.g., changing Some to All) before syllogistic reasoning was initiated this would most likely lead to a flawed judgment albeit one that did not arise from a logical misstep.

c. One of the premises may have been omitted from the reasoning process. Arising prior to the reasoning process, this omission could lead to the mistaken conclusion that logic itself, failed.

d. An additional premise may have been incorporated into the reasoning process.

Interestingly, Henle suggested that what she terms “pre-logical errors,” may be more likely for syllogisms that contain implausible and/or emotionally loaded content; as though such material discourages analytical processing from the get-go. Such content may be a reason for the participant to reject the reasoning task, add or omit premises, or otherwise tamper with the task.

The Figural Effect

Frase (1968) reported that participants rapidly evaluated Figure 1 as opposed to Figure 4 syllogisms (see Table 1).

Table 1

*Figure 1 and Figure 4 Syllogistic Forms*

<table>
<thead>
<tr>
<th>Figure 1:</th>
<th>Figure 4:</th>
</tr>
</thead>
<tbody>
<tr>
<td>All B are C (B-C)</td>
<td>All C are B (C-B)</td>
</tr>
<tr>
<td>All A are B (A-B)</td>
<td>All B are A (B-A)</td>
</tr>
<tr>
<td>All A are C (A- -C) matches (A-B-C)?</td>
<td>All A are C (A- -C) matches (C-B-A)?</td>
</tr>
</tbody>
</table>
Fraser applied an association learning paradigm to explain these results, reasoning that Figure 1 is easier for participants to evaluate because such a structure involves **forward chain association learning** (B-C then A-B = A-B-C) whereas Figure 4 necessitates **reverse chain association learning** (C-B then B-A = C-B-A which is the reverse of A-C).

**Elaborating the Role of Structural Variables**

Seven core hypotheses have been proposed to explain why syllogistic reasoning often ends with a flawed validity judgment. Some of these hypotheses center on the role of structural variables whereas others emphasize syllogistic content. The goal of this coming section is to trace the path of researchers who followed up on the structurally-oriented hypotheses before eventually turning to a consideration of more contemporary works that have come out of the content-oriented tradition established by Janis and Frick (1943). What can be concluded based on more recent tests of the atmosphere, conversion, and figure hypotheses? Endeavoring as I am, to merge the structural and content traditions, the key findings about the role of structure must be considered as I design my belief bias studies. In this way the complementary approach I intend to pursue, as well as my incorporation of novel structural variables, will be accomplished in a manner that hopefully sheds light on the key questions being pursued in the field of syllogistic reasoning research.

**Which is Dominant? Atmosphere or Conversion?**

The atmosphere effect proposed by Woodworth and Sells (1935) was actively refuted as an explanation for syllogistic reasoning errors by Chapman and Chapman (1959) who offered the conversion effect as a key factor underlying flawed syllogistic reasoning. Simpson and Johnson (1966), Begg and Denny (1969), Dickstein (1975) and finally Revlis (1975) pitted these hypotheses against each other. Simpson and Johnson as well as Begg and Denny’s work
provided support for the atmosphere effect whereas Revlis’ later research in 1975 led to the contemporary perspective that atmosphere and conversion both impact reasoning but that it is the conversion effect that reasoners must first avoid if they are going to render an accurate judgment.

Simpson and Johnson (1966) reasoned that if atmosphere accounted for syllogistic reasoning errors then offering ‘anti-atmosphere training’ should reduce said errors. Secondly, that if this training failed to reduce syllogistic errors, it might be concluded that errors arise from some other source, perhaps conversion. And finally, that conversion could be tested in a similar manner, by offering ‘anti-conversion training’ and observing its effect on subsequent syllogistic reasoning. Procedure-wise, participants underwent ‘anti-atmosphere’ or ‘anti-conversion’ training which consisted of a description of the relative effect and an opportunity to complete 6 practice syllogisms (it should be noted that this second component of training was only a part of ‘anti-atmosphere’ training, an inconsistency that is addressed shortly as a key weakness of this study). Simpson and Johnson found that ‘anti-atmosphere training’ reduced errors but that ‘anti-conversion training’ did not have this effect. Based on these results the authors concluded that the atmosphere effect, rather than conversion, was at the base of syllogistic reasoning errors.

I would like to highlight several weaknesses associated with Simpson and Johnson’s (1966) study that I feel impose limitations on what can safely be concluded from this work. The between-subjects design of this study compared several groups, most notably those who received ‘anti-atmosphere’ and ‘anti-conversion’ training. These groups underwent different training and also evaluated different sets of syllogisms. Between-group variation in accuracy might therefore be attributed to stimulus-based differences as opposed to training differences; specifically, the set of syllogisms that was administered to the ‘anti-conversion’ group may have been more difficult than those administered to the ‘anti-atmosphere’ group. As well, the ‘anti-atmosphere’ group
received practice trials while the anti-conversion group did not. These reasons alone could have translated into poorer relative performance post-training for the anti-conversion group.

Compared to Simpson and Johnson’s (1966) questionable approach, Begg and Denny (1969) addressed the atmosphere-conversion conflict in a manner that was arguably more rigorous. Begg and Denny provided evidence that the atmosphere and conversion hypotheses are not irreconcilably different. Specifically, both hypotheses predicted the same errors for most syllogism types. The exception was for syllogisms starting with IE, EO, and OE premises; in these cases the hypothesized errors were different depending on whether one considered the potential effect of atmosphere or conversion. For instance, with IE syllogisms (Major premise: some are; Minor premise: none are), the atmosphere hypothesis would lead to the prediction that participants’ responses would favour conclusions of the form O (Some are not), whereas the conversion hypothesis would predict that conclusions of the form E (None are) would be most likely. Begg and Denny collected new data in an empirical attempt to reconcile the interpretations for syllogistic reasoning errors. They employed a wider diversity of invalid syllogisms and found data that corroborated Woodworth and Sells’ atmosphere-based account of syllogistic reasoning errors.

Dickstein (1975) pursued Simpson and Johnson’s (1966) line of reasoning and offered up an improved methodological means for testing the purported protective effect of ‘anti-atmosphere’ training as well as a re-evaluation of ‘anti-conversion’ training. In particular, the same set of syllogisms was administered to participants in the ‘anti-atmosphere’ and ‘anti-conversion’ groups; both groups had received comparable training and were allowed to reference their training materials during the test trials. Dickstein reported that improvements in reasoning were most marked when training guarded against conversion as opposed to atmosphere. In fact,
the ‘anti-atmosphere’ training condition was not found to improve performance much above the level observed in the study’s control condition.

These findings directly challenged Simpson and Johnson’s (1966) as well as Begg and Denny’s (1975) conclusion that atmosphere was the dominant factor, but it is important to note that Dickstein’s findings may derive from a study design that had its own weaknesses. Upon reviewing the instructions provided in the ‘anti-atmosphere’ and ‘anti-conversion’ conditions, it becomes apparent that the former condition may have been less successful because the instructions were more difficult to apply to the syllogistic reasoning task. In order to avoid bias derived from the atmosphere effect, participants must obey a number of rules that guide the evaluation of three propositions at once, whereas bias associated with conversion can be avoided with relatively greater ease, by applying a single rule to one proposition at a time. Thus, although each training regimen provided participants with detailed information about a particular bias, the design of the study did not ensure that participants would have the same capacity to override these biases between conditions.

Revlis (1975) elaborated the atmosphere (Woodworth & Sells, 1935) and conversion (Chapman & Chapman, 1959) theories so that they made definite predictions about participants’ performance for all syllogism types as opposed to a smaller set of syllogisms like the ones that had been described by Begg and Denny (1969). Revlis then documented participants’ errors on a set of 64 syllogisms and calculated the predictive efficacy of the atmosphere and conversion models that he had elaborated. Revlis replicated previous work that used a limited set of syllogisms, showing that the atmosphere and conversion models were equally effective predictors. However, upon comparing the two models for the remainder of the previously
untested syllogisms, the atmosphere model was 71.4% accurate as a predictor whereas the conversion model was substantially less successful.

Despite these findings, Revlis proceeded to discount the efficacy of the atmosphere model because he said that it could not account for patterns of accurate reasoning that were demonstrated by participants who were predicted to reason inaccurately for certain syllogism types. Revlis concluded that the best model would be one that presumed that the conversion and atmosphere effects work in tandem. Specifically, participants who illegally converted the premises and came to the conclusion that none of the provided answers were correct, would take a second pass through the premises, and at that point be likely to fall prey to the atmosphere effect. Revlis’s perspective was the one that I adopted for my own research. Given that the conversion and atmosphere effects are documented sources of syllogistic reasoning error and that I am interested in studying belief bias effect as a source of error in syllogistic reasoning, I attempted to exclude syllogistic forms that would make the atmosphere effect likely to occur. Secondly, I ensured that the likelihood of the conversion effect was controlled for across the conditions of my studies.

*What about Figure? Does it Factor in as a Possible Source of Syllogistic Reasoning Errors?*

A review of the literature to this point has identified two key structural variables that must be incorporated into the design of syllogistic reasoning studies. A third structural variable that was encountered in the literature and that stands out as a potential confounding factor if not controlled would be figure, or *mood* as it is often called. Frase (1968) described how Figure 1 and Figure 4 were resolved with relatively more or less success and attributed these differences to the direction of association learning that was necessitated during each trial type. Frase based his conclusions about the figural effect on trials that were non-semantic. Interestingly, it seems
that adding semantic content to syllogisms activates a second set of figural effects that
distinguish Figures 2 and 3 (see Table 2) based on accuracy (Pezzoli & Frase, 1968).

Table 2

*Figure 2 and Figure 3 Syllogistic Forms*

<table>
<thead>
<tr>
<th>Figure 2:</th>
<th>Figure 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>All C are B (C-B)</td>
<td>All B are C (B-C)</td>
</tr>
<tr>
<td>All A are B (A-B)</td>
<td>All B are A (B-A)</td>
</tr>
<tr>
<td>All A are C (A-C)</td>
<td>All A are C (A-C)</td>
</tr>
<tr>
<td>matches (C-B-A)?</td>
<td>matches (A-B-C)?</td>
</tr>
</tbody>
</table>

Frase (1968) was initially the one to hypothesize that participants would be more accurate
when reasoning about Figure 2 as opposed to Figure 3 premises because Figure 2 premises
elaborate on the conclusion terms (C and A) whereas Figure 3 premises do not elaborate the
conclusion terms and instead provide competing information about an intermediary term (B).
Despite this rationale, Frase (1968) did not find a significant difference between Figures 2 and 3.
Pezzoli and Frase (1968) suggested that this was because non-semantic syllogisms were used by
Frase (1968) and that such syllogisms were unlikely to produce different levels of *interference*, a
content-dependent factor that would be likely to inhibit reasoning for Figure 3 syllogisms.

Pezzoli and Frase (1968) manipulated the degree of semantic association (none, low, high) between the terms in the major and minor premises as well as the amount of verbiage (few words, many words) in the argument to determine if this would differentially impact accuracy for Figure 2 and Figure 3 syllogisms with an invalid form. Semantic association was defined according to Bilodeau and Howell’s (1965) associative norms (as cited by Pezzoli & Frase, 1968). More specifically, semantic association within a syllogism was deemed to be entirely *absent* when the terms were abstract nonsense syllables; *low* when there was less than a 10% chance that the three terms would be correlated (e.g., teapot, haystack, astronaut); and high when
there was more than a 10% chance that the three terms would be correlated (e.g., needle, haystack, farmer). Participants’ ability to deduce that syllogisms were invalid was no different according to figural form for the no association and low association conditions which replicates Frase’s (1968) finding for non-semantic syllogisms. A figural effect was observed for syllogisms that contained a high degree of semantic association; specifically participants were more likely to be inaccurate when judging Figure 3 syllogisms that contained terms sharing a high degree of relatedness than on Figure 2 syllogisms with comparable content.

In a nutshell, participants were more likely to mistake invalid syllogisms for valid ones when the risk of semantic interference was high, as is the case for arguments that elaborate about a peripheral, non-conclusion term. In order to judge the conclusion of a Figure 3 syllogism, which is a statement about A and C, participants must inhibit what they have been told in the previous premises about B. Thus, to know whether or not “All farmers own a sewing needle” (All A are C) follows logically from the premises that “All haystacks contain a sewing needle” (All B are C) and “All haystacks are owned by farmers” (All B are A), one must in essence inhibit what they know about haystacks (B) and orient their judgment towards farmers (A) and sewing needles (C). It is relatively difficult to inhibit what is known about haystacks since they are semantically associated with farmers and sewing needles. Interference was documented as being less severe when the terms within the syllogism shared little in common semantics-wise. So, it is less difficult to determine whether the conclusion “All astronauts own a teapot” (All A are C) follows logically from the major premise “All haystacks contain a teapot” (All B are C) and minor premise “All haystacks are owned by astronauts” (All B are A), because one can more easily break the syllogism into its components when those parts are semantically unrelated.
In the case of Figure 2 syllogisms, participants’ judgments were equally successful regardless of whether the content was low or high in terms of the degree of association between terms. Presumably, even if there is a high degree of semantic association between terms, participants are not going to face interference from the premises when evaluating the conclusion. This is because the information in the premises in Figure 2 contributes directly to one’s evaluation of the conclusion; premises 1 and 2 provide central information about A and C, which is the conclusion itself.

Roberge (1971) questioned the methodological rigour of the studies conducted by Frase (1968) and Pezzoli and Frase (1968). Roberge argued that the atmosphere effect might account for the error patterns that, according to Frase and Pezzoli, resulted from a figural effect or a structure (figure) x content (degree of association) interaction. Roberge tested his hypothesis by administering a large set of syllogisms where the atmosphere of the premises was rarely a cue for the accurate conclusion. The rare occasions when atmosphere was likely to bias participants’ judgments were balanced across the figural conditions. Despite these methodological adjustments, the key finding from Frase’s (1968) study was replicated. Figure 4 syllogisms involving a reverse chain of processing were found to be more difficult to solve than Figure 1 syllogisms involving a forward chain of processing. Figures 2 and 3 were found to be resolved with equal accuracy which, given that Roberge’s stimuli were non-semantic, is a finding that is consistent with Frase’s (1968) reported null effect and Pezzoli and Frase’s (1968) finding for neutral and low semantic association trials.

Johnson-Laird and Steedman (1978) followed up on Frase’s (1968) initial finding that Figure 1 syllogisms are easier to solve than Figure 4 syllogisms, presumably because “forward chain” learning is easier than “backward chain” learning. Using a wider variety of valid and
invalid syllogisms than in Frase’s original study, Johnson-Laird and Steedman concluded that it is not that Figure 1 is easier than Figure 4, but rather that a figural bias is at work. Specifically, Figure 1 biases reasoners to pick A-C conclusions whereas Figure 4 biases reasoners to pick C-A conclusions. Given that Frase (1968) had reasoners evaluating the validity of syllogisms ending exclusively in A-C form, a full figural effect was not found, and instead Figure 1 was assumed to be easier than Figure 4.

Johnson-Laird and Steedman (1978) went on to propose a new theory of syllogistic reasoning since the paired-associates learning paradigm advanced by Frase (1968) was deemed irrelevant given the figural bias that had been detected. Johnson-Laird and Steedman proposed an analogical process of syllogistic reasoning consisting of four stages: (a) semantic interpretation of the premises, (b) heuristic combination of the premises, (c) formulation of a tentative conclusion in line with the heuristically combined premises, and (d) a logical testing of the tentative conclusion which may lead to its alteration or abandonment. Accordingly, there are heuristics or shortcuts that tend to guide how these premises are combined. For instance, people have a throughgoing bias, that is, a bias toward linking up end (1st and 3rd) items by way of middle items… thereby settling on the conclusion that “Some artists are chemists” rather than “some chemists are artists,” “some artists are not chemists,” or “some chemists are not artists”.

Bucci (1978) extended Johnson-Laird and Steedman’s work by applying a Piagetian model to explain the analogical process underlying syllogistic reasoning. Accordingly, syllogistic reasoning is accurate and logical when the reasoner is able to apply grammatical rules in their surveying of (a) a universal statement, and (b) a three part syllogistic argument. Bucci reported that children were less successful than adults when asked to construct buildings out of blocks according to universal premises like “all the yellow blocks should be squares”. To accurately
complete this task participants needed to distinguish the subject (yellow blocks; square blocks) from the predicate (squares; yellow objects) rather than viewing them as one entity. This would permit them freedom when building, for instance “all the yellow blocks should be squares” – if parsed without grammar limits the reasoner to constructing a building with only yellow, square blocks; if parsed grammatically affords the reasoner freedom to construct a building with yellow square blocks as well as square blocks that are of other colours. Adults succeeded at this type of task which is to say that they were successful at applying grammatical rules when interpreting universal statements.

The second part of Bucci’s study looked at the type of semantic relationship shared between the subject and predicate. A broad predicate was one that applies to the subject but also to other categories (e.g., All dogs are mammals – it is easy to generate a non-dog exemplar for the mammal category). A narrow predicate was one that applies mostly to the subject and less often to other categories (e.g., All flightless birds are penguins – it is more difficult to think of a non-penguin exemplar for the flightless bird category). Finally, the abstract predicate was one that is less likely to be considered in a real-world context (e.g., All the large blocks are orange). Bucci points out that being able to generate a real-world representation of the premises may help reasoners apply grammatical rules. In terms of accuracy when evaluating syllogisms adults were substantially better at solving syllogisms starting with a major premise containing a broad predicate than a narrow one. The narrow predicate condition was more accurately resolved than the abstract predicate condition. Children showed a similar pattern although their overall accuracy was substantially lower than that of the adults. Bucci suggested that this is because children merely judged the arguments based on their ability to generate a counter-example (thus if they couldn’t think of another flightless bird they would be biased by this)
whereas adults used their everyday knowledge to generate counter-examples but also as a way of making salient to themselves the grammatical structure of the argument – “the evocation of a specific referent may aid assignment of sentence structure by establishing the fundamental distinction between reference and attribution” (Bucci, 1978, p. 20).

*Summary: Structural Variables*

Atmosphere, conversion and figure are three key factors, derived from syllogistic structure, that impact syllogistic reasoning. The biases that occur as a consequence of these factors must be considered and controlled for within my own studies. As is highlighted by Bucci’s work, structure is not independent of content, and so we turn our attention to the relationship between these variables in Chapter 2.
CHAPTER 2:
REASSESSING THE RELATIONSHIP BETWEEN STRUCTURE AND CONTENT

The propositions in a syllogism may assert “facts” about the world that conflict with observable features of said world. So too may the propositions induce emotions in the reasoner. In both instances, the semantic content of a syllogistic argument is known to impact reasoning and it is a primary goal of my dissertation to understand this process. A proper elaboration of the role of content variables necessarily takes place against the backdrop of what has been concluded in the past by reasoning researchers. Reasoning researchers have traditionally emphasized reasoning over emotions and structural variables over content variables. I would argue that this perspective originates in part from the philosophical reference points to which these researchers were exposed and less so from their inquiries into the nature of reasoning itself. As such it is important to understand the potential biases in the field that may have contributed to the way that content variables, specifically beliefs and emotions, have been construed and tested.

The Historical Polarity of Structure and Content

How has the interaction between structural variables and content been treated historically in the literature? A review of the literature shows that structural variables have been celebrated whereas non-structural variables (including content factors) have been treated as the dishonest cousins of logical form (Falmagne & Gonsalves, 1995). The presumed lack of parity between errors derived from structure and those derived from content is an assumption that I would like to avoid in my own work because I do not feel that it reflects an unbiased analysis.

Historically speaking there are many illustrations of the assumption that structure and content are not on par. Structural factors were linked with a special brand of cognition, one that was not entangled in the trappings of everyday life, a process that was deemed ‘pure’ and
insulated. The syllogism was viewed as “a tool of logical analysis rather than a diagram of any typical reasoning process,” (Woodworth & Sells, 1935, p. 451). Logical inference has been fêted as an activity for educated minds and perhaps less accessible to the ‘common’ mind, “the principles of logical inference are techniques which are not the common property of the unsophisticated subject and must be distinguished from psychological inferences which may be made by the ordinary person” (Lefford, 1946, p. 144).

“Good errors” (Köhler, 1927, as cited in Henle, 1962, p. 369) result from roadblocks that have defeated “objective and clear thinking” (Lefford, 1946, p. 146) whereas “stupid errors” (Köhler, 1927, ibid.) are the “onslaughts of outrageous fortune” (Lefford, ibid.), and “solutions obtained through blind procedures” (Wertheimer, 1959, as cited in Henle, 1962, p. 369). Content factors including beliefs and emotion are the notorious opponents of clean deduction. A reasoner can be felled by their beliefs; “the discovery of truth is prevented more effectively, not by the false appearance things present and which mislead into error,” (read: whether the atmosphere and figure afford an appearance of validity), “not directly by weakness of the reasoning powers,” (read: how well one is trained in formal logic), “but by preconceived opinion, by prejudice” (Schopenhauer, 1788-1860). Clearly, the consensus amongst early reasoning scholars was that content is a formidable foe of syllogistic reasoning. In the realm of everyday life, where calculated actions are valued, the content of arguments is deemed all the more powerful in that emotions induced by content can lead to inaccurate conclusions about one’s circumstances and unwise actions,

“The disastrous effects of emotional thinking can be found everywhere from demagogy in politics to unfortunate personal adjustments in living. The problem is especially acute today, in a war-torn world, where only action based on objectivity of analysis and rationality of thought can lead to a successful solution of the social and economic problems which gave rise to it and consequences which will result from it.” (Lefford, 1946, p. 127)
Reasoning scholars have traditionally devalued non-structural variables and contended that logical deduction is an objective process that *should* rely only on the structural features of the syllogism. Although this perspective may reflect empirical findings, namely that structure plays a dominant role in reasoning and therefore should be emphasized in research; I would argue that this perspective also stems from broader philosophical trends. Firstly, reasoning scholars’ perspectives accord with and may have been influenced by the New Critics’ emphasis of “close reading” and objective paths to understanding, a perspective that dominated American literary studies during the 1920s through 1950s. And secondly, reasoning scholars’ preference for cold and cognitive considerations of structure over intuitive reckonings of content overlaps, perhaps by more than mere coincidence, with a cognitive ‘obsession’ that dominated the history of research in Western psychology and entirely discouraged the development of emotion-oriented research until the mid-1980s when this branch of psychology finally began to emerge as its own dominant force. If reasoning scholars’ treatment of structural and content variables has been shaped by other than the scientific process then, the role of content variables may have been unjustifiably diminished, and the contention that content disrupts or disengages the reasoning process, may not have been adequately tested.

**New Criticism**

New Criticism developed in the early decades of the 20th century, eventually becoming the dominant approach to literary theory in the 1950s (Harth, 1981). It seems likely that the New Critics’ treatment of the poem penetrated other disciplines, for instance impacting how logicians treated the logical syllogism. Some evidence to support this claim includes the fact that moral education, which prioritized logical training, was common in North America throughout the first half of the 20th century (Benjafield, 2010). This would have ensured that the fundamental
principles underlying deductive reasoning and logic were familiar to those in the educated classes, many of whom may have gone on to pursue formal studies in logic and/or literary theory. Logicians and literary theorists were often one and the same, for instance Bertrand Russell was a scholar of philosophy, logic, mathematics and history (Benjafield, 2010).

The New Critics believed that a poem is *self-referential* and *self-contained* and that a successful reader is one who is able to filter out external factors while surveying the poem. Poetic understanding was seen as the product of an objective process that excluded societal and historical context, the author’s intentions, and the reader’s experiences and emotions. Proponents of New Criticism redefined poetry; “expressive criteria of spontaneity and overt emotion… of the Romanic lyric” were replaced by “language and structure” (Harth, 1981, p. 526). This new definition of poetry and poetic process idealized structure and shunned emotion, “poetry is not a turning loose of emotion, but an escape from emotion; it is not the expression of personality, but an escape from personality” (Eliot, 1919, as cited in Harth, 1981, p. 528). Given the New Critics’ conviction that “the exclusive criteria for poetry were no longer drawn from the creative mind or sensibility behind the work but from the verbal structures constituting the works themselves” (Harth, 1981, p. 526), “personal emotion in poetry” was what led works by the likes of John Dryden and others to be written off as “unpoetic” (Harth, 1981, p. 527).

The assumed primacy of objective structural elements over potentially biasing effects of emotional content and context was a hallmark of New Criticism. Yet, this distinction coincided with reasoning theories that similarly differentiated objective structure from subjective content. Although it is possible that reasoning theorists came to the same conclusions as their literary counterparts by way of an independent, empirical path of inquiry, it is also possible that logicians
may have favoured structure over content (as well as reasoning over emotions) because this practice was *de rigueur* at the time in neighboring academic fields.

*History of Psychology*

Falmagne and Gonsalves (1995) linked cognitive psychology’s conceptualization of reasoning as an insulated, technical process with rationalist philosophy which described deduction as “encapsulated” (p. 1) and held that abstract and contextual elements were natural opponents. They argued that reasoning is a complex process that is under-estimated by such structural definitions, that a comprehensive understanding of deduction relies on the holistic study of “formal and contentful factors” (p. 1), and that current debates within cognitive psychology about how best to account for reasoning are unproductive because they polarize as opposed to integrate factors associated with the process.

In line with a tendency to de-value the role of content, the Behaviorist and Cognitive eras in Western psychology have also traditionally de-emphasized emotion, acknowledging it in turn as either an ‘unknowable force’ or as a mere accessory of cognition (Lazarus, 1999). Emotions tended to be defined as bodily or non-cognitive, as “vivid feelings rather than forms of thought” (Dixon, 2012, p. 340), a conceptualization that explicitly polarized emotion and cognition, leading scholars to study cognition by observing the non-emotional subject (Dixon, 2012) as opposed to the subject who thinks and feels simultaneously. In the 1950s Magda Arnold and Sylvan Tomkins contended that the discipline of psychology was neglecting the study of the emotions and called for a reinvigoration of this research area (Keltner, Oatley & Jenkins, 2013). Arnold and Tomkins may have increased the visibility of emotions; however they continued to be treated as accessory factors playing second fiddle to cognition for several decades to come. Only in the 1980s did scholars such as Zajonc (1980) challenge the presumed primacy of
cognition by providing evidence of affective judgments being independent of and preceding in time, the cognitions that normally were assumed to precede them. Works such as these encouraged a new wave of emotion-centered research to emerge in the late 1980s (Lazarus, 1999).

As noted, a cognitively-oriented paradigm dominated North American psychology between the 1920s and 1980s. The psychological study of syllogistic reasoning in this period similarly tested cognitive theories of reasoning and avoided investigating the role of emotionally-driven processing of syllogisms. Upon questioning what the basis was for researchers to pursue a cognitive as opposed to cognitive-emotional or emotional program of syllogistic reasoning research, one might conclude that researchers were exclusively, empirically driven. However, it seems unlikely that researchers were entirely unswayed by the broader paradigm and instead chose to test theories of syllogistic reasoning in which cognition was given primacy over emotion. This would have ensured that their research was consistent with contemporary perspectives in psychology; however it would not have ensured that cognition and emotion were treated in a way that reflected their actual roles in a syllogistic reasoning task. As Blanchette and Richards (2010) contended, “higher cognitive processes were studied in a vacuum, separately from the affective system, as if they were immune from such influence. This partition may have stemmed from early conceptual distinctions between reason and passion, with its implicit hierarchical distinction.” (p. 561)

An Attempted Rapprochement of Structure and Content

A preliminary critique of the history of reasoning research leads me to suspect that the conclusions of reasoning scholars about the role of content variables like beliefs and emotions may in part reflect the pro-structure philosophical zeitgeist of their times as well as the anti-
emotion paradigm that predominated in Western psychology from the 1920s to the mid-1980s (Lazarus, 1999). Given this historical context, it seems that speculation about the relative role of structure and content must continue.

Content in a syllogism that leads to belief bias and/or emotions is not necessarily less important than structure and I would contend that it does not always disengage one from the pursuit of a logical train of thought. Certainly there are circumstances where beliefs and emotions make it so that reasoning is less likely, for instance, as per Köhler, Thorndike’s cat was too frustrated to make neither heads nor tails of his confinement in a puzzle box, for “to place a starving kitten in the cramped confinement of one of Mr. Thorndike’s box-cages, would be more likely to make a cat swear than to lead it to act rationally” (as cited in Benjafield, 2010). Henle (1962) similarly suggested that the content of an argument may be in such conflict with one’s beliefs, or arouse such a level of emotion, that any intention to engage in syllogistic reasoning is derailed from the beginning. However, to contend that belief bias and emotional experience are necessary boundaries that always prevent syllogistic reasoning is indeed an abrupt conclusion.

In turning our discussion specifically towards belief bias I would hypothesize that a conflict between our beliefs and the content of a syllogism does not turn off reasoning but rather changes it. In line with this idea is the assumption that reasoning is not a clear-cut process where structural variables are more potent than content, or vice versa. I do not assume that there is a separation between content and structure but rather am attempting a rapprochement between the two, an interaction which is aptly described by Margery Allingham (1889-1996), a British mystery writer. She contemplates that a conclusion may arise, not “by the decent process of quiet, logical deduction, nor yet by the blinding flash of glorious intuition, but by the shoddy, untidy process halfway between the two” (as cited by Bailey, 2007, p.143).
Elaborating the Role of Content Variables

Establishing that Belief Bias is a Genuine Phenomenon

Janis and Frick (1943) documented what they termed a belief bias effect when they observed that a conflict between the plausibility of content and validity of structure led to errors in syllogistic judgment. A common criticism of Janis and Frick’s study was that their effect might be accounted for by atmosphere. Morgan and Morton (1944) responded to this criticism by explicitly pitting atmosphere and beliefs against each other. The authors designed a set of non-semantic syllogisms and documented the effects of atmosphere. They then inserted content into these frames, producing a set of syllogisms where the belief bias effect would be predicted to act in direct opposition to the atmosphere effect. For instance, one syllogism that was administered by Morgan and Morton had a conclusion, “X may not be more effective than Z,” which was favoured by atmosphere in its non-semantic form, yet discouraged by beliefs, in its semantic form because its content, “All airplanes (X) may not be more effective than battleships (Z),” was not in line with the popular conviction in the 1940s that airplanes were more effective than battleships. Morgan and Morton reported that in this situation where atmosphere and beliefs were in opposition, participants’ judgments of syllogistic form betrayed belief bias more often than bias derived from atmosphere.

Conditions that Inspire Belief Bias

The following sections describe factors that have been shown to cause errors in reasoning. Some of these studies have explicitly looked at the process by which such errors are produced, implicating belief bias directly, whereas others have focused more on the fact that errors have occurred without engaging in a direct discussion of belief bias itself. For instance, numerous authors conclude that certain conditions prompt the application of heuristic strategies,
of which belief bias is one. As such, to ensure the broadest possible consideration of antecedents for belief bias, studies that address belief bias directly, as well as those whose findings have indirect implications related to belief bias, will be described. These factors include: emotional content in the argument, felt emotion in the reasoner, strains on working memory capacity and executive functioning, other variables associated with the reasoner (e.g., attitude strength, reasoning abilities), and finally other variables associated with the argument (e.g., evaluation or generation task, difficulty level, source credibility).

Emotional content. Lefford (1946) reported that emotional syllogisms are more apt to inspire belief-biased judgments than non-emotional syllogisms. This finding is not to suggest that neutral content fails to inspire belief bias but rather, that emotionally evocative content is more successful as a source of bias. That said, critical reflection about this study brings to light some methodological issues that might limit the confidence with which one can make any conclusions about Lefford’s findings. Specifically, Lefford did not verify that the emotionality and plausibility manipulations employed in his study were effective. So, despite some findings that shed light on the possible antecedents of belief bias, the key take-home message from this study is a methodological one pertaining to pre-testing and the verification of manipulations.

Addressing much the same question as Lefford, Blanchette and Richards (2004) assessed the effects of positive, negative, and neutral content on participants’ propositional reasoning about statements of the form “if p, then q”. Much like a logical syllogism, the problem’s two parts may be neutral or non-emotional (e.g., if one skis, then one burns calories versus if one skis, then one may get injured), and may be followed by a valid or invalid conclusion. In Experiment 1 participants were more likely to answer illogically when evaluating arguments embedded with anxiety-, sadness-, and happiness-related words than neutral ones. In Experiment
2 a stricter paradigm was used to ensure that the previously observed difference between the neutral and emotional conditions was not due to word-related factors other than valence. Thus, neutral words were transformed into emotional words through classical conditioning, ensuring that the treatment and control groups encountered the same words that differed only in their associated valences. Blanchette and Richards replicated the earlier effect, which was that emotional content led to more errors, regardless of valence. What is notable about this second experiment is that conditioning yielded a set of “emotional” words that were relatively less emotionally intense than the ones used in Experiment 1. Pairing “sandwich” with 5 negative images yielded a conditioned stimulus that was unpleasant but not nearly as negative as an unconditioned word like “death”; even so, it was as successful at impeding logical performance which shows that even subtle emotionality in an argument can derail rationality. Blanchette and Richards offered two interpretations of these findings. Firstly, that emotionality impacts working memory, specifically by putting greater demand on working memory resources than neutral materials. Or secondly, that emotionality impacts participants’ interpretation of conditional statements such that these types of statements are inaccurately assumed to function biconditionally (“if p, then q” is assumed to imply “if q, then p”).

Working with a stimulus set comprised of “emotional” syllogisms that were created through classical conditioning or the embedding of intrinsically unpleasant words within the arguments, Blanchette and Leese (2011) replicated the deleterious effect of negative emotional content on syllogistic reasoning accuracy relative to neutral argument content. Blanchette and Leese extended their analysis of negative emotion beyond that of previous studies by comparing the relative importance of participants’ implicit physiological arousal and their explicit evaluation of syllogistic content to reasoning accuracy. They reported that it is in fact the level of
implicit arousal (as measured by skin conductance) that seems to impact accuracy as opposed to one’s explicit awareness of an argument’s emotionality (assessed through stimulus ratings). Specifically, a high level of implicit arousal coincided with logically inaccurate evaluations of syllogisms containing negative content, an effect that was not observed when participants explicitly expressed that they found the stimuli to be highly emotional.

The difference between implicit arousal and explicit evaluation is one that needs to be highlighted here because it may in fact be that the interaction between the implicit-explicit dimension and arousal-evaluation processes is what determines one’s reasoning success. Blanchette and Leese (2011) found that implicit arousal hurts reasoning while explicit awareness of the argument’s emotionality does not. We can speculate that belief bias may be stronger when the conclusion represents an unconscious position as opposed to one that is salient, because what is implicit cannot be inhibited. This conclusion seems reasonable based on classic work by Schwarz and Clore (1983) in which participants’ estimates of personal well-being were only colored by their weather-induced moods when these emotional states were implicit and had not been consciously reflected upon (as cited in Keltner et al., 2013).

*Emotional frame of mind.* The previous group of studies addressed the question of how emotional content in a logical problem may impact reasoning. Emotion may also impact reasoning when it is induced in and felt by reasoners who then are asked to evaluate neutral arguments. Getting at this question, Melton (1995) worked from the observation that positive moods are often associated with worse performance on analytic tasks and provided evidence to suggest that errors in syllogistic reasoning by happy participants are due to an increased application of heuristic strategies, of which belief bias could be one. Melton noted that this effect might be explained in two ways - positive moods may distract or discourage energy expenditure
by participants on the task at hand. Melton expected that longer response times would be observed in distracted participants and shorter ones would occur in the case of reduced effort. Additionally, Melton expected that reduced effort would translate into more frequent use of heuristic strategies (e.g., atmosphere), less use of diagrams, and more confidence related to syllogistic judgment because affect was being used implicitly as information.

Participants who had undergone a positive mood induction were more likely to render inaccurate judgments and pick atmosphere-congruent conclusions, more frequently chose universal as compared to qualified (some) conclusions, drew fewer diagrams, and used slightly less time on each trial than control participants. These findings supported the reduced effort hypothesis leading Melton to speculate that mood could be acting in two ways to reduce effort – through mood maintenance or affect as information. Specifically, happy participants may not have felt that solving syllogisms was a ‘fun’ activity and therefore may have been unmotivated to partake in a mood-disrupting task. Alternately, participants’ positive mood may have led them to feel confident that they have reached an accurate judgment sooner than if they were in a neutral or negative mood. In either case, this study suggests that when generalized to a discussion of belief bias, positive mood may be a factor that makes it more likely to occur.

Oaksford, Morris, Grainger and Williams (1996) compared facilitation and suppression hypotheses to explain the relationship between mood and reasoning. The former expectation was derived from the finding that positive mood helps divergent thinking and creativity, presumably by making long-term memories more accessible and encouraging better quality working memory processes (Isen et al., 1987, as cited by Oaksford et al.), thus reasoning should be improved when participants are feeling positive. The suppression hypothesis was based on the resource allocation theory of depression – that participants in negative moods retrieve negative extra-task
or task-irrelevant information, which is taken to imply that negative and positive moods will function similarly by making it likely that participants will be distracted by negative or positive memories, respectively, that are not important for the task at hand.

Experiment 1 used a derivative of the Wason card selection task in which falsification would be a logical strategy and confirmation would be non-logical. Participants in positive and negative induced moods sought confirmation more often than falsification, a finding that supports the suppression hypothesis. Experiment 2 attempted to directly assess the impact of working memory on reasoning by supplementing the card selection task with a concurrent monitoring task that would put strain on working memory. In this case the key between-subjects comparison was not mood related, but rather degree of working memory strain. Oaksford et al. reported that a non-logical confirmation strategy was more common when participants were completing a concurrent task than not, which suggests that working memory load encourages non-logical processing.

Finally, in an effort to directly test the relationship between mood and working memory, the Tower of London task (Shallice, 1982, as cited by Oaksford et al., 1996) was employed in a third experiment in which participants were asked to complete an activity that requires planning and comparison of various strategies in working memory. This task was assumed to tap executive functioning and mood was hypothesized to lead to extended response times because when the central executive is not working optimally, participants have a harder time deciding which plan they should go with. Alternately, if response times were comparable to those of controls, it was predicted that participants would opt for a less efficient plan. It was reported that in positive moods, participants took the same amount of time to pick a plan compared to controls, yet this plan was less efficient because it required extra moves. In contrast, participants
in negative moods took longer than controls but in the end opted for an efficient plan that was comparable to the one selected by controls. Together these results suggest that positive mood decreases performance on tasks requiring planful behaviors controlled by the executive functioning system, whereas negative mood did not seem to compromise planning behavior expect in the case of limited response times.

*Emotional content and emotional reasoners.* Previous studies of reasoning assessed the effects of emotion on reasoning by applying an orthogonal strategy whereby emotional content and emotional state of the reasoner were independently manipulated and there was little connection between syllogistic semantics and participants’ affect. These types of studies are said to be investigating *incidental affect* which is an emotional state in the reasoner that is not related to the semantic content of the reasoning task (Blanchette & Richards, 2010). In contrast, Blanchette, Richards, Melnyk, and Lavda (2007) presented emotional syllogisms that were explicitly, semantically related to participants’ naturally-occurring affective states, a design that permitted them to study *integral affect* which is an emotional state that is linked to the contents of the reasoning task (Blanchette & Richards, 2010). This difference in the types of affect under investigation may account for the fact that Blanchette et al.’s (2007) findings countered the general consensus of previous studies – that emotions impede accuracy in reasoning tasks. In particular, Blanchette, Richards, Melnyk and Lavda (2007) studied individuals who had witnessed shocking acts of terrorism in an effort to determine whether their reasoning about generally emotional and terrorism-specific emotional topics would be more or less impeded by belief bias than their reasoning about neutral topics. Blanchette et al. compared three groups of participants: individuals who had directly witnessed a bombing in London, UK on July 7, 2005; individuals living in Manchester, UK who witnessed footage of the event on television but had
no direct experience with it; and Canadians from London, Ontario who had minimal to no knowledge about the event.

Interestingly, participants from London, UK reported the highest levels of fear and were the most accurate at evaluating terrorism-related syllogisms in which content and validity conflicted. Participants in Ontario reported the least fear and were the least accurate on these types of syllogisms. Secondly, reasoning accuracy on these syllogisms was positively correlated with a risk estimate measure employed in the study, specifically, participants who made the fewest reasoning errors on terrorism-related incongruent trials also estimated that there was a low likelihood of a similar attack occurring in Britain in the next month. Together these findings led the authors to conclude that proximity to terrorism may lead to increased feelings of fear, yet an ability to inhibit terrorism-related beliefs (as evidenced by low risk estimates) helps these same participants reason accurately about potentially biasing content. A follow-up study 6 months later showed that Londoners continued to feel greater fear than other participants, and also reported greater anger, a difference that did not exist when first assessed. Despite this new emotional backdrop to reasoning, Londoners performance on terrorism-related emotional trials continued to be superior to that of the other groups.

Much like Blanchette et al. (2007), Vroling and de Jong (2009) investigated belief bias in relation to integral affect (fear about negative social evaluation). However, whereas Blanchette et al. reported that feeling fear diminished belief bias reasoning about terrorism-related syllogisms, Vroling and de Jong found that fear was associated with greater likelihood of falling prey to belief-confirming material in social anxiety-related syllogisms. More specifically, Vroling and de Jong showed that individuals suffering from social anxiety disorder have a greater tendency than control participants to accept as true conclusions that confirm a fundamental, negative belief that
they hold about themselves and/or social situations (e.g., ‘I am unpopular’). Biased thinking of this type is said to be the main challenge faced by individuals experiencing social anxiety; specifically those with elevated levels of fear about negative social evaluation tend to give less weight to disconfirmatory evidence that would challenge negative views of the self (e.g., writing off a positive social experience) than information that helps to “maintain fearful preoccupations” (p. 634). What is important to note is that this style of reasoning was restricted to syllogisms about social themes and that a confirmatory bias was not more likely for socially anxious participants than controls when faced with neutral content.

The discrepancy in these findings may be accounted for in several ways. For one, Blanchette et al. documented that participants who gave low risk estimates for the possibility of another terrorist act in London were the least prone to belief bias, a finding that led them to conclude that effective coping relies on one’s ability to control beliefs tied to negative emotional events. In Vroling and de Jong’s study, there is no way of knowing how much control socially anxious participants had over negative beliefs associated with social situations, nor how well they were coping. If the socially anxious participants in Vroling and de Jong’s study were not in control of their beliefs and were not coping with their social anxiety through belief inhibition, then Vroling and de Jong’s findings do not refute Blanchette et al.’s explanation. Secondly, the conditions in both studies were not comparable to each other. In Blanchette et al.’s study, terrorism-related syllogisms were highly emotional for Londoners exposed to terrorism directly, and moderately emotional for the other two groups of participants. Yet, in Vroling and de Jong’s study, social anxiety-related syllogisms were highly emotional for the socially anxious, but presumably minimally emotional or even non-emotional for their socially secure counterparts. Thus, finding a significant difference in reasoning about social anxiety-related syllogisms
between the socially anxious and non-socially anxious may be a product of this stimulus-level
difference and not a finding that conflicts with Blanchette et al.’s principal finding about how
individuals reason about material that is either highly emotional, or moderately so.

In a review about the relationship between affect and reasoning, Blanchette and Richards
(2010) remarked on a divide in the literature. While some researchers had documented a negative
effect of emotions on reasoning which they tied to emotion’s demanding draw on working
memory resources, other researchers reported that emotion was associated with improved
reasoning possibly because it helped to promote belief inhibition. Blanchette and Richards
elaborated the theory behind the second position, noting that the working memory account could
not explain emotion’s occasional role as a reasoning facilitator. Instead of highlighting working
memory, they pointed to attention as the mechanism of interest. They contended that incidental
emotion (felt but not integrated in the task) encourages the consideration of task-irrelevant
information; whereas integral emotion (felt and integrated in the task) focuses attention on task-
relevant information (and in the case of Blanchette et al.’s (2007) Londoners, helps with the
inhibition of task-irrelevant, terrorism-related beliefs which would be useless given the
propositions at hand). At the same time, Blanchette and Richards concluded that because
evidence does exist for both a memory and attention mechanism, future research should aim to
integrate the roles of working memory and attention into an explanation of how reasoning
unfolds under emotional circumstances.

Goel and Vartanian (2011) provided additional support for Blanchette and Richards’
contention that an attention-level (as opposed to a working-memory level) mechanism may be at
work when reasoners solve syllogisms more accurately under emotional circumstances.
Although previous studies had documented belief bias in experiments where emotionality or
believability were manipulated separately, Goel and Vartanian (2011) manipulated both of these variables simultaneously to allow for an examination of the interaction between them when it comes to reasoning errors. They demonstrated that participants were less likely to commit errors in accordance with beliefs on emotional as compared to neutral trials. Specifically, participants were better able to detect invalid structures despite the appeal of believable conclusions, and secondly, participants took longer to accurately judge these invalid trials, implying greater deliberation. Goel and Vartanian referred to Forgas’ (1995) Affect Infusion Model to explain these effects. This model proposes that negative emotions encourage more vigilant and systematic consideration of an argument’s propositions, a view that highlights the role of attention in the reasoning process.

In a follow-up paper, Vartanian and colleagues (2013) showed that negative syllogistic content can impact participants’ performance on syllogistic reasoning trials in the response phase without impacting the shape of reasoning itself. They drew from recent memory studies that have revealed that emotions increase participants’ subjective feeling of recollection but do not enhance actual memory recall accuracy. Vartanian et al. concluded that the negative syllogistic content in their study did not impact reasoning because participants’ accuracy and reaction time were equal on incongruent trials regardless of whether the arguments contained negative or neutral conclusions. In line with memory research, Vartanian and colleagues provided evidence to suggest that a response-level bias might be at work. This bias seemed to be characterized by the tendency to accept invalid and implausible arguments as ‘valid’ more often and with more confidence when they contained negative rather than neutral content. This type of response represents an attenuation of belief bias which normally manifests itself in the form of extremely low rates of acceptance for invalid-implausible trials.
Elaborating on their results, Vartanian and colleagues speculated that negative content may produce a shift in responding, leading to a repositioning of the response criterion so that it becomes more liberal. In other words, participants lower their expectations about what a “valid” argument is like and are less stringent than they would normally be when comparing arguments containing negative content to this model or criterion. This means that they are more likely to conclude that the argument is valid. This response level bias is thought to be a product of a fast and heuristic type of processing, whereas the reasoning stage effect of negative affect demonstrated by Blanchette et al. (2007) and Goel and Vartanian (2011) implicates a slow and deliberate processing system. What may determine whether belief bias impacts reasoning or responding is the intensity of emotion evoked by the syllogism. Low-intensity negativity may have been experienced in the current study and not have been potent enough to impact reasoning per se whereas the negative emotions evoked by Blanchette and Goel and Vartanian may have been stronger and more consequential. This is an interpretation that could account for the difference, particularly if we note that participants were recruited from within the Department of National Defence. Presumably these individuals are exposed to negative themes like insurgency and mortality more than the average person because of their workplace.

*Working memory and executive functioning.* The picture that became clear in the previous section is that a further factor that may prompt belief-biased responding relates to working memory and executive function. Specifically, if participants are cognitively overburdened, perhaps due to age-related memory deficits, multi-tasking or there are excessive task demands, they may not apply a logical strategy and instead resort to a heuristic strategy such as consulting the argument’s atmosphere or content. Alternately participants may possess the logical know-
how and requisite memory resources but due to a failure in executive functioning, be unable to ‘turn on’ their logical thinking cap or prioritize logical skills over heuristics.

We can start by reviewing evidence that shows how working memory may at times fail, lead to reasoning errors that may have resulted from belief-biased evaluations. For one, compared to younger participants, older participants have been shown to demonstrate substantially greater rates of belief bias when reasoning about neutral content (e.g., “all juicy fruits are oranges”) (Gilinsky & Ehrlich, 1989). This finding was replicated by Gilinsky and Judd (1994) who noted that older participants showed an exaggerated acceptance bias (accepting as true something that accords with one’s beliefs) which could have resulted from heuristic processing or alternately from older individuals generally having “strongly entrenched belief systems” and the desire to “preserve stereotypes and resist modifications” (p. 370). Gilinsky and Judd tested three alternate hypotheses, each of which could explain why heuristics might have been dominant in older adults. They found evidence that aging leads to a decline in working memory resources and vocabulary comprehension, factors that independently predicted low accuracy and which could have inspired a heuristic style of problem-solving. They also found that aging might make it more difficult to handle complexity, evidenced by older participants having particular difficulty when asked to solve complex 3-model syllogisms compared to the simpler 1-model variety. In a later paper by De Neys and Van Gelder (2009) it was reported that relative to young adults, children and older adults, performed poorly on conflict trials in which beliefs and logic were incongruent. This poor performance was thought to stem from low belief inhibition capacity, a decontextualization process that is itself manifested as a curvilinear trend across the lifespan.
Besides age-related factors, mood seems to also impact working memory. Channon and Baker (1993) completed a study that did not explicitly examine belief bias; however their key finding supports the notion that heuristics (such as belief-guided processing) are more common when people are experiencing memory and/or executive functioning deficits. Working with a non-semantic set of syllogisms, Channon and Baker (1993) investigated this exact possibility by comparing the primary syllogism-solving strategies used by depressed participants to controls. Their study was inspired by research that has demonstrated that depressive deficits seem to interfere with effortful processing by either diminishing the availability of problem-solving resources or reducing an individual’s ability to initiate cognitive problem-solving strategies that were not automatically prompted by the task itself (e.g., Hasher & Zacks, 1979; Hertel & Hardin, 1990, as cited in Channon & Baker). Channon and Baker reported that compared to controls, the depressed problem-solvers were least likely to use logic and most often opted for a simple heuristic strategy relying on one rule. Although the depressed participants might have applied a heuristic strategy based on atmosphere, the authors noted that this was not particularly common which suggests that applying two rules to complete an atmosphere-guided judgment (Woodworth & Sells, 1935) may not have been possible because participants may have forgotten one of the two premises. Thus, applying a strategy where one had merely to recall one of two premises was the best strategy for depressed participants who were presumably strapped for resources in working memory. If Channon and Baker had employed semantic syllogisms I would expect that their findings would translate into greater belief-biased processing in the depressed group because such a strategy relies on the application of a single rule – determining if one part of the argument – the conclusion – is empirically true.
Moutier and colleagues (2006) highlighted the role of executive functioning in syllogistic reasoning. They demonstrated that it is a failure in executive functioning, specifically a lack of inhibitory control, that leads to belief bias and errors in reasoning. In other words, participants may have the logical skills and the memory resources necessary to accurately judge a syllogism but fail to do so because they fail to inhibit a non-logical strategy suggested to them by the contents of the syllogism. For instance, an unbelievable conclusion may suggest that unbelievable-equals-invalid is an appropriate strategy; yet this strategy is illogical and needs to be inhibited if one hopes to accurately affirm a valid syllogism as such. Moutier and colleagues derived evidence for this conclusion from observations made during the application of a negative priming paradigm aimed at determining whether inhibition is required for accurate syllogistic judgments. They reasoned and went on to show that participants have difficulty affirming a previously inhibited conclusion, something that leads to errors in the evaluation of valid syllogisms. This may be due to the fact that participants inhibited the real-world corollary (“all elephants are heavy”) of an implausible syllogistic conclusion (“all elephants are light”) in order to correctly affirm an argument’s validity. This application of inhibitory control was successful on trials in which real-world beliefs conflicted with validity, yet on ensuing trials this inhibition impeded performance because participants were required to affirm the validity of an argument ending with the previously inhibited statement (“all elephants are heavy”).

*Other personal qualities of the reasoner.* Gilinsky and Judd (1994) demonstrated that individual differences associated with aging, specifically low tolerance for complexity and poor vocabulary, often precede belief bias. Feather (1964) examined other individual differences that may contribute to belief bias. He tested three hypotheses about the antecedents of belief bias, specifically that belief bias will be greater when i. *attitude intensity* and ii. *intolerance for...*
inconsistency are high, and when iii. logical abilities are low. Male participants evaluated neutral, pro-religious, and anti-religious syllogistic arguments. Participants’ scores on the neutral semantic trials served as an index of logical ability and their degree of agreement/disagreement with the syllogistic content was measured and used to distinguish attitude intensity. Finally, a measure by Budner (1962) which assessed ‘intolerance to ambiguity’ was administered in an effort to measure tolerance for inconsistency (as cited in Feather, 1964). Feather calculated the degree of intercorrelation between the three dependent variables and found support for all three of his hypotheses when considering the pro-religious participants (majority of his sample). Interestingly, his hypotheses did not hold for the anti-religious group but this may be due to the fact that there were very few of them and that they were only mildly anti-religious. Pennycook, Cheyne, Koehler, and Fugelsang (2013) have done more recent work exploring the role of religiosity, religious skepticism, and one’s vulnerability to belief bias.

Argument variables. Markovits and Nantel (1989) identified two additional factors that make belief bias particularly likely – syllogistic production tasks (as compared to evaluation tasks) and difficulty. Interestingly, the importance of this first variable came to light because Markovits and Nantel used a new method to address Henle’s (1962) concerns that participants sometimes inadequately attend to the premises. They compared two types of syllogistic tasks: the first being the traditional evaluation task in which participants were seen as likely to skip over the premises; and the second being a novel production task in which participants were asked to generate the conclusion rather than merely evaluating it against the backdrop of the premises. Markovits and Nantel hypothesized that participants would be less impacted by their beliefs if asked to produce a conclusion because they would not skip to the conclusion and be distracted by their beliefs about its plausibility. The key findings from this study include that participants
appeared to be more biased by their beliefs when asked to produce a conclusion, a result that countered Markovits and Nantel’s hypothesis; and that the likelihood that participants would commit an error due to belief bias related to problem difficulty. General difficulty level of the problems was measured by looking at the error rates on neutral trials of a particular syllogism type. The neutral trials with the greatest number of errors were the ones where belief bias was most likely to run rampant when a belief-validity conflict was created in the conclusion. On the other hand, problems with easy structures were less likely to be solved inaccurately when embedded with incongruent content in the conclusion.

*Source credibility* is another variable associated with the content in a syllogism that contributes to belief bias (Copeland, Gunawan & Bies-Hernandez, 2011). As Copeland and colleagues noted, source credibility can depend on the honesty or the expertise attributed to a source. As such, two experiments were completed in order to determine the relative effects of each factor. Participants were instructed to verify a set of syllogisms in which the conclusions were described as inferences that had been made by honest versus dishonest (Experiment 1); or experienced versus novice (Experiment 2) sources. It was hypothesized that heuristic, belief-first processing would be more likely when participants trusted the source because they were portrayed as honest or experienced. On the other hand, elaborated logical processing was expected in cases where the source was distrusted due to a dishonest or novice portrayal. In Experiment 1, it was found that for invalid syllogisms, participants committed more errors when they trusted the supposedly honest source. A similar and significant pattern of errors was found in Experiment 2 which suggests that source credibility, be it the product of perceived honesty or expertise, promotes belief-first, heuristic processing.
Why Belief Bias Happens

A key investigation of the process underlying belief bias was conducted by Oakhill, Johnson-Laird and Garnham (1989) who offered a two-part explanation. Firstly, that one’s beliefs about syllogistic content can impact the likelihood of conversion (e.g., all X are Y is more vulnerable than All spaniels are dogs which is semantically blocked from being converted). And secondly, that in cases where conversion is not possible but reasoning errors still occur, beliefs must be the culprit in that they discourage the construction and evaluation of alternative models. According to this explanation, if the conclusion is plausible a reasoner may not scrutinize alternate models whereas if the conclusion is implausible this may promote the construction and review of alternate mental models.

A comparison of one-model valid syllogisms revealed that participants were more likely to accept as true arguments with plausible as opposed to implausible conclusions. Invalid syllogisms were also included as stimuli and Oakhill and colleagues found that false positives (accepting as true a false conclusion), were more likely when the conclusion was plausible. Most importantly, beyond documenting the traditional belief bias effect, Oakhill and colleagues also observed that when scrutinizing an invalid syllogism, participants were most likely to reject the conclusion and go on to generate a conclusion that would render the syllogism valid when that conclusion contained implausible content. This suggests that participants search for alternative models more when they are faced with incongruity between problem content and everyday knowledge than when these things match. These results also suggest that implausible content makes it easier to detect invalid structure and to remedy this structure, rendering it valid.

It should be noted that Oakhill, Johnson-Laird and Garnham (1989) also incorporated valid syllogisms with multi-model solutions as stimuli in their study. They hypothesized that
implausible content in the conclusion of valid syllogisms would push participants to look for alternative models which would translate into high accuracy for solving valid, multi-model syllogisms. Yet, Oakhill and colleagues found nothing more than a non-significant pattern that approximated their suspicions leading them to conclude that even though implausible content might have promoted a search for alternative models, this search would not have been fulfilled because multi-model syllogisms are too difficult and it is extremely challenging to construct or to examine an array of alternative models. Thus, the motivation to search for alternatives is present but the will or way is absent. Thus, the difficulty level of syllogisms may be a limiting factor when it comes to the possible effects of beliefs on reasoning.

Wyer and Goldberg (1970) derived an explanation for syllogistic reasoning from the analysis of probabilistic reasoning. Although their treatment of syllogistic reasoning does not explicitly discuss belief bias, their model could account for it, as I will argue. Objective probability rules outline the equation $P_{AB} = P_A \times P_B$, which in essence means that the probability that A and B will occur equals the product of the probability that A and the probability of B occurring. This logic can be translated into the syllogistic task by replacing the phrase “probability that A will occur” with that of “probability that A is an accurate statement about the world”. Assuming that participants are looking at each premise and judging its plausibility on a scale of probability (highly unlikely through highly likely), the equation becomes $P_{\text{conclusion}} = P_{\text{premise1}} \times P_{\text{premise2}}$.

If an individual is following objective probabilistic reasoning rules in their evaluation of semantic content in syllogisms, they should afford the same weight to each of the premises. Additionally, they should recognize that the probability of a conclusion hinges on the two premises such that $a$: 
a. highly probable conclusion results from two highly probable premises;
b. highly improbable conclusion results from two highly improbable premises; and
c. moderately probable conclusion results from either two moderately probable premises or one highly probable and one highly improbable premise.

Wyer and Goldberg compared the expected and observed probability estimates that participants gave for premises and conclusions and found that generally the rules above were followed. So, the probability of a conclusion was roughly the average of the probabilities of premises 1 and 2. According to Wyer and Goldberg’s approach, reasoners faced with a neutral semantic syllogism afford equal weights to the parts of the argument and as such engage in a logical process of deduction where the conclusion is assumed to be worth no more than each of the premises. Because the conclusion has been deemed highly probable based on probabilistic reasoning, the reasoner engages in a syllogistic task in which the rules of syllogistic reasoning are maintained – the parts of the argument are equal. When the conclusion is deemed to be improbable, Wyer and Goldberg’s explanation suggests that probabilistic reasoning cues participants to violate syllogistic reasoning rules by presuming that some parts are truer than others. Thus, according to Wyer and Goldberg’s perspective, belief bias may be due to an unequal weighting of the parts of the argument - either the conclusion is mistrusted and afforded less power, or one of the premises is seen as less important than the other.

Pragmatic normalization is another explanation that has been used to account for the process underlying belief bias. Fillenbaum (1974) asked participants to paraphrase messages that don’t fit with real-world knowledge (e.g., “If you don’t stop I won’t sue you”; or “If you don’t hurry up you’ll catch the bus”). Although a syllogistic paradigm was not being examined, the key point from Fillenbaum’s study nonetheless sheds light on a process that may be the foundation for belief bias in syllogistic tasks, specifically, participants tended to normalize the meaning of the sentences so that the message would fit with pragmatic real-world scenarios. This
shows that when people are merely asked to encode sentences (i.e., they are not performing a memory recall task), they tend to adjust the content so that it fits with pragmatic knowledge of an actual world. Translated into the syllogistic reasoning paradigm, Fillenbaum’s finding suggests that upon encountering a conflict between real-world knowledge and syllogistic content, participants may adjust the content to diminish the conflict before engaging in syllogistic reasoning. Thus, belief bias may be due to inaccurate encoding of the argument’s components.

Two major challenges to this account of belief bias include for one, that belief bias is still observed when participants are asked to generate the conclusion, something you would not expect to find if belief bias resulted from a mismanagement of a given conclusion statement. And secondly, that the observed accuracy patterns associated with the four key belief bias conditions (valid-believable; valid-unbelievable; invalid-believable; invalid-unbelievable) cannot be accounted for by the adjustments one would predict if pragmatic normalization were to precede syllogistic reasoning. For instance, a valid syllogism with an implausible conclusion ought to be encoded such that the conclusion becomes plausible and the conflict is resolved pre-logically. This adjustment would necessarily boost accuracy, however what is found in this classic condition is rather that accuracy plummets when the conclusion of a valid syllogism is implausible. This suggests that pragmatic normalization has not occurred.

Reaching a Consensus by the Eighties?

Janis and Frick (1943) first documented belief bias in action within syllogistic tasks. Their work was followed by the work of others who replicated this effect (e.g., Morgan & Morton, 1944), identified conditions that would make it more likely to occur (e.g., Lefford, 1946; Feather, 1964; Markovits and Nantel, 1989), and proposed why it might take place (e.g., Wyer & Goldberg; Fillenbaum, 1974; Oakhill, Johnson-Laird & Garnham, 1989). Despite the evidence
garnered from these studies to substantiate the existence of the belief bias effect, there were still some researchers in the eighties who denied that it was a key factor behind syllogistic reasoning errors. Revlin, Leirer, Yopp and Yopp (1980) were skeptical about the importance of belief bias. They did not deny that such an effect exists but described it as a ‘dishonest’ error. Following in a tradition that was established by Chapman and Chapman in 1959 and pursued by in the seventies by Dickstein and Revlis, Revlin et al. focused on the role played by conversion, a bias that was deemed to be an ‘honest’ source of error.

Revlin et al. (1980) designed two experiments in which they attempted to find evidence that would support their Rationalist position and be consistent with the view that errors in syllogistic reasoning are due to illegal conversion of the premises. In Experiment 1, participants decided which non-controversial conclusion (including “no conclusion possible”) followed from two controversial premises. Conversion (assuming that premise “All A are B” implies that “All B are A”) can lead to errors in syllogistic reasoning when it causes the reasoner to pick a conclusion drawn from the illegally represented premises, yet sometimes the reasoner gets lucky and the conclusion is the same regardless of the illegal conversion. Revlin et al. (1980) found evidence to support the conversion theory in that participants’ accuracy was higher when conversion coincidentally implied the same conclusion than when conversion implied a different conclusion.

Revlin et al. reasoned that if conversion leads to errors in syllogistic reasoning when a different conclusion is implied then anything that can be done to block conversion will improve reasoning and provide further evidence for the conversion theory. The authors noted that real-world knowledge can block conversion (e.g., knowing that “All dogs are mammals” does not imply that “All mammals are dogs”). Participants in Experiment 1 indicated how likely they
would be to endorse the converse of premises in the study and it was reported that “nonconverse believers” – those whose access to real-world knowledge blocked conversion – were better at evaluating syllogisms in which illegal conversion would have led to endorsement of an inaccurate conclusion. Revlin et al. accepted this as evidence for conversion, a “logical error,” being at the root of syllogistic reasoning problems yet they did not acknowledge that beliefs and real-world knowledge were necessary antecedent conditions for this process of non-conversion. The fact that beliefs can block conversion was noted by Oakhill, Johnson-Laird and Garnham (1989) who offered an entirely different interpretation, one that highlighted rather than discounted the belief bias effect.

Experiment 2, as conducted by Revlin and colleagues, looked at the interaction between believability and validity of the conclusion statements of arguments. They reported that congruent syllogisms were solved more accurately than incongruent ones, a finding that was commonly found and at the core of belief bias theories. Yet, Revlin et al. discounted this finding by highlighting possible methodological confounds that could be at the root of this pattern in their data. For one thing, Revlin et al. observed that in situations where beliefs about content were in conflict with validity and errors were observed, the opted-for conclusion was not always one that accorded with beliefs (e.g., the most believable option). Rather, when conflict was observed and participants had to choose the conclusion from a multiple choice list they often picked “No conclusion is proven” which is essentially inconclusive. Revlin et al. suggested that this indicates that belief-validity conflicts do not lead to an abandonment of logic and the application of heuristically-driven, biased evaluation but rather, that participants are doing the logical thing by selecting an inconclusive answer. I would argue that the design of Revlin et al.’s study doesn’t adequately pressure participants for an answer, nor does it adequately contrast
believability and validity. If participants could not select the inconclusive response what would they do? Most likely they would demonstrate a more robust belief-biased response. Instead, when confronted with an implausible conclusion participants opt to make no decision – they do not engage in logical reasoning that would lead them to the accurate conclusion, nor do they engage in belief-biased judgment. If reasoners were actually rational, belief-logic conflict would not lead to an inconclusive judgment. Instead, reasoners would render a judgment that was logically derived as though the semantic content were irrelevant.

Although Revlin et al. replicated key findings reported by proponents of the belief bias effect; their interpretations clearly disagree with those who argue that belief bias is a key antecedent for syllogistic reasoning errors. A critical survey of the assumptions and methods used by Revlin et al. has led me to conclude that Revlin et al. underestimated the role of belief bias. This suspicion on my part is supported by an article written by Evans, Barston, and Pollard (1983) who addressed the debate between so-called Rationalists (e.g., Revlin et al.) and Anti-rationalists (e.g., Oakhill et al.) by critically assessing the literature and conducting three experiments in which past methodological shortcomings were corrected for.

In brief, the Rationalist position supposes that reasoners form abstract representations of syllogistic propositions and apply a set of inferential rules to said models. A logical error therefore must result from atmosphere and/or conversion effects. Beliefs are not conceived of as a source of logical bias because reasoners deal only with structure. Reasoners who deal with beliefs and content are not reasoners because they are thinking non-logically, and therefore any error that they make is a non-logical one. In contrast, the anti-Rationalist position emphasizes the role of problem content and semantic associations and acknowledges that these problem features impact reasoning and can cause legal errors.
The first part of Evans et al.’s article highlighted limitations that were common to both perspectives, including the reporting of weak effects, poorly worded questions, a failure to control for confounds such as atmosphere and conversion, and carry-over effects resulting from the repeated use of the same participants. Evans et al. then went on to critique Revlin et al.’s (1980) work, noting that there were three reasons why their study would have been insensitive to belief bias effects. For one thing, the multiple choice method applied by Revlin et al. was seen as a limitation because while Revlin et al. claimed that they were offering four possible conclusions of which one was “plausible,” a closer examination of the conclusions revealed that two out of four counted as “plausible” options. Thus, participants may have been picking a belief biased response twice as often as Revlin et al.’s measure of bias let on. Evans et al. also argued that Revlin et al.’s predominant use of valid syllogisms and lack of control of atmosphere across conditions were additional limitations. As Evans et al. discussed, it is possible that belief bias occurs more often in the evaluation of invalid syllogisms, and that atmosphere is a strong effect that may mask or distort the occurrence of belief bias unless prevented or controlled for.

The second part of Evans et al.’s article was a report of three experiments that tested for belief bias having eliminated the problems noted in previous research. The three experiments documented a belief bias effect and provided evidence for the conclusion that this effect, while occurring in valid trials, is somewhat more likely in invalid ones and that participants are more sensitive to logical validity on unbelievable trials, as though plausibility conveys a false sense of security. Evans et al. provided an important foundation for later syllogistic reasoning research by validating the role of argument content and reporting a key interaction between beliefs and validity that later models must be able to explain.
CHAPTER 3: SYLLOGISTIC REASONING MODELS

Working from the assumption that a sufficient explanation of syllogistic reasoning incorporates not only structural factors but also provides an explanation of belief bias, researchers in the eighties and onward have developed a variety of syllogistic reasoning models.

One key question that these models address, relative to belief bias, is whether it occurs because deductive reasoning is *interrupted* by the plausibility of the conclusion or because the reasoning process is *modulated* (rather than inhibited) given the plausibility of syllogistic content. A second fundamental question that these models address is how reasoners attend to the parts of the argument – whether they engage in a forward style of reasoning in which premises are represented prior to the conclusion, or whether reasoning takes place in a reverse order where the conclusion takes precedence and the premises are considered belatedly. The goal of the coming sections is to summarize and compare the six key models before turning to the research that I completed. My studies were in many ways able to replicate conventional effects reported in the literature; however the addition of new variables led to some findings that cannot be accounted for by any of the current models. The six key models are:

i. Selective scrutiny model (Evans, Barston, & Pollard, 1983).


iv. Verbal reasoning theory (Polk & Newell, 1995; Thompson et al., 2003)

v. Selective processing account (Klauer, Musch & Naumer, 2000).

vi. ROC account (Dube, Rotello & Heit, 2011).
The Selective Scrutiny Model

Evans, Barston, and Pollard (1983) proposed the so-called selective scrutiny model, an account that explains why there is a larger effect of belief bias in invalid, believable trials than in valid, unbelievable trials. In brief, this model suggests that participants evaluate the plausibility of a conclusion as their first step in the syllogistic judgment task. If the conclusion is believable they tend to judge that the problem is valid without engaging in deductive reasoning. If the conclusion is unbelievable it is at this point that participants engage in deductive reasoning and are more apt to render an accurate judgment. Essentially, \textit{logical deduction is said to be driven by disbelief}. This explanation accords with observations made by Lord, Ross, and Lepper (1979) who reported that participants were more critical of contradictory than confirmatory evidence. Interestingly, Lord, Ross and Lepper’s participants did not merely \textit{dismiss} contradictory evidence (essentially ignoring it), but rather expended considerable critical energy in an attempt to dismantle said evidence. Lord, Ross and Lepper termed this form of processing \textit{biased assimilation}, which applies nicely to the syllogistic paradigm in which participants assimilate or integrate the parts of a syllogism, at times in a belief biased manner.

The Misinterpreted Necessity Model

An alternative model follows from a second explanation offered by Evans et al. (1983) which centered on the fact that participants may experience uncertainty when examining indeterminate syllogisms and therefore be more vulnerable to beliefs because of the fact that the premises do not necessitate nor deny a particular conclusion in some syllogisms. Markovits and Nantel (1989) formalized this interpretation as the \textit{misinterpreted necessity} model, having employed production as opposed to evaluation tasks in their study of syllogistic reasoning. They found that participants were more likely to display belief bias on invalid trials and speculated
that structural uncertainty associated with the indeterminate nature of some invalid (but no valid) syllogisms is what pushes participants to switch over to a non-logical strategy. They assumed that participants start out by generating potential conclusions based on the premises and then abandon this logical strategy when it leads to an indeterminate conclusion. Thus, a belief-oriented, “second-order strategy… [is] used in order to accommodate a degree of internal conflict or uncertainty” (p. 16), or as Dube et al. (2010) put it, “belief-biased responding [is] an escape-hatch mechanism when deductive reasoning is inconclusive” (p. 833). This approach is termed *misinterpreted necessity* because it assumes that reasoners are uncertain about what they should render as a judgment when they encounter a conclusion that is *possible* but not *necessitated* by the premises. It is important to note that the original formulation of this model by Markovits and Nantel did not deny (nor did it concede) the occurrence of uncertainty and belief bias for valid and determinately invalid syllogisms. Addressing this issue, Evans and Pollard (1990) and Newstead, Pollard, Evans and Allen (1992) developed a caveat to the original (ironically indeterminate!) model by exploring the possibility that reasoners can feel uncertain about valid and determinately invalid syllogisms, particularly when problems are complex and difficulty level is high. Thus, although belief bias is more common in indeterminate syllogisms, it may also be observed when reasoners hit a road-block in their logical assessment of other types of syllogisms.

**The Mental Models Account**

Despite providing a caveat that would help to extend the misinterpreted necessity explanation, Newstead et al. (1992) provided empirical support for the mental models theory which comes out of work by Oakhill, Johnson-Laird & Garnham (1989). The mental models account of syllogistic reasoning itself derives from the broader mental models framework
detailed by Johnson-Laird (1983). The mental models account supposes that there are three consecutive stages in the deductive reasoning process: comprehension, description, and validation. In brief, this theory suggests that reasoners construct a mental representation of the premises (the comprehension stage) by encoding the parts of an argument as visual tokens. In this way the statement “all dogs are mammals” would be encoded by envisioning tokens representing dogs which all fall within the mammal category. Secondly, reasoners are presumed to generate a tentative conclusion (or compare a given conclusion) to the model (the description stage) that is implied by the premises. Finally, this account posits that reasoners test this tentative conclusion by attempting to construct alternative models that would falsify it (the validation stage). Johnson-Laird and Byrne (1991) have argued that it is only in the validation stage that true critical analysis occurs, “only in the third stage is any essential deductive work carried out: the first two stages are merely normal processes of comprehension and description” (p. 36) (as cited in Polk & Newell, 1995).

Comprehension and description are assumed to take place before a belated consideration of the conclusion’s plausibility occurs. The validation stage, the point of true critical analysis, is thought to be more likely to occur when the conclusion is implausible. Accordingly, if a conclusion fits the envisioned framework and is plausible, it is likely to be accepted without undergoing validation. If however the conclusion fits but is implausible, the reasoner is highly likely to engage in a search for alternatives that would falsify it. This model is similar to the selective scrutiny model in that participants are said to engage in a logical search for alternative conclusions after finding that a conclusion is implausible. The key difference is that in the selective scrutiny model an individual only reasons if the conclusion is implausible (implying that no logic is applied in the evaluation of syllogisms with believable conclusions). In contrast,
according to the mental models approach, an individual firstly engages in logical reasoning about the premises and conclusion, then as a second step evaluates the conclusion’s plausibility. They *only re-appraise* the conclusion by constructing alternative models if the conclusion is unrealistic. Thus, the evaluation of plausibility is a first step in the selective scrutiny model and either a final or interim step according to the mental models perspective.

The mental models approach suggests that two factors determine how difficult a syllogism is – the figure of the premises (which makes it more or less difficult to construct these propositions and produce a conclusion) and the number of mental models that one needs to construct in order to test this conclusion. Having to construct and compare many mental models puts an obvious demand on working memory and in many cases it may be impossible to consider all possible models, and rather verification will take place based on a limited consideration of possible models. This of course makes it likely that the critical model will not be considered during verification, thereby leading to errors in reasoning.

Evidence to support the mental models theory was reported by Oakhill et al. (1989) who compared the rate of belief bias for three types of syllogisms embedded with plausible or implausible content: (a) one-model valid syllogisms; (b) indeterminate syllogisms; and (c) multiple model valid syllogisms. While the results for (a) and (b) illustrated their model, (c) was associated with inconclusive findings which the authors suggested were due to the inordinate difficulty level of multiple model syllogisms. Specifically, participants evaluating one-model valid syllogisms were more likely to accept as true conclusions that were plausible rather than implausible. This supports mental models theory because it shows that knowledge blocks participants’ acceptance of implausible content thereby prompting a search for an alternative solution that does not exist. When the search for an alternative fails the participant rules that
there is no valid conclusion because the original unbelievable (yet valid) option has already been dismissed. Thus, beliefs operate at two loci – they lead to a sometimes fruitless search for alternatives and rule out any further consideration of an option earlier deemed to be implausible. For invalid indeterminate syllogisms, participants were more likely to accept them in error when they were believable and correctly detect their invalid nature when they were unbelievable.

Oakhill et al. (1989) expected that implausible content in the conclusion of valid multiple model syllogisms would push participants to look for alternative models, thereby making them better at solving valid syllogisms that have multiple models. Yet, they did not find this and concluded that even though implausible content may set off alarm bells and promote a search for alternative models, this search may be fruitless because there are too many alternative models to juggle and compare. Thus, a motivation to search may be present yet the will/way can be absent.

Verbal Reasoning Theory

Polk and Newell (1995) proposed the verbal reasoning theory to account for deduction. They based their theory on the assumption that people’s linguistic skills are more developed than any reasoning-specific skills they might possess and therefore that an average person, when faced with a logical syllogism, will apply their more sophisticated skill set, one that taps their knowledge of syntax. Although reasoning-specific skills are acknowledged to exist, Polk and Newell contend that they are relatively under-developed compared to linguistic skills and therefore are less likely to be applied. Accordingly, receiving extensive training in logic would develop reasoning-specific skills, making for a specialized set of abilities that might rival one’s linguistic tool box.

Their theory represents an extension of Lance Rips’ (1983) argument that deduction is a verbal process whereby syntactic rules are used to generate inferences. Polk and Newell
compared their theory with that of mental models (Johnson-Laird, 1983) and agreed that comprehension, description and validation occur during reasoning. Yet, instead of being represented as visual tokens that come to be integrated through the envisioning of diagrams and figures, Polk and Newell proposed that the parts of arguments are represented as verbal units that come to be integrated through the application of syntactical rules that govern how parts in speech relate to each other (e.g., subject versus object designation). Additionally, Polk and Newell stated that “true reasoning” is not restricted to the validation stage but rather, that comprehension and description are central to deduction.

Polk and Newell (1995) provided evidence for their verbal reasoning theory by referring to a computational model in which a computer program (rule-based linguistic manipulation of terms) reliably solved categorical syllogisms. To perform the comprehension stage this computer program relied on annotation that distinguished between is (e.g., a painter is a sculptor = painter'sculptor) and is not (e.g., a painter is not a sculptor = painter—sculptor). The description stage relies on a generate-and-test procedure whereby the program creates conclusions that could be true based on the annotated premises previously constructed and then tests each conclusion to determine if its structure is legal (i.e., does it relate two end terms and does it apply one of the four standard syllogistic moods?) If a proposed conclusion is legal it is accepted by the program. If it is illegal then the program moves into the third phase in which it attempts to generate alternate conclusions. The key to this re-encoding phase is that the program considers explicit knowledge derived directly from the annotations (e.g., some painters’sculptors) but it also takes into consideration indirect knowledge that is implied by the annotations, yet implicit in the problem (e.g., some painters’sculptors implies that the situation some painters—sculptors is
possible although it is not explicitly stated). This third phase is essentially an effort to extract more knowledge, which may lead to a solution.

The role of content and beliefs was incorporated into the computational model by creating annotations that specified empirically plausible and implausible propositions. Whenever the model generated an implausible conclusion it automatically entered the third phase and attempted to extract more information to construct a conclusion that would accord with both empirical and structural rules. The belief bias effect observed in humans was replicated by the computational model by instructing the program to make a random choice between “not valid” and a conclusion based around beliefs if this re-encoding process failed to yield a logically and empirically true conclusion.

An adjustment to Polk and Newell’s verbal reasoning theory was put forth by a Canadian group of researchers at the University of Saskatchewan, Thompson et al. (2003), who noted a discrepancy between observed response times and those predicted based on Polk and Newell’s conceptualization of verbal reasoning. Many explanations of belief bias, including the mental models and verbal reasoning theories by Oakhill et al. (1989) and Polk and Newell (1995) respectively, assume that reasoners are more critical when conclusions are implausible than when they accord with empirical knowledge. From this assumption one can derive expectations about how long reasoners will spend thinking about a given type of proposition.

Hypothesis 1 is therefore that reasoners will spend more time evaluating arguments ending with an implausible conclusion than a plausible one. When validity and believability interact and the entire process of syllogistic evaluation is considered, clear hypotheses can again be made about the rate of responding. Hypothesis 2 is that reasoners will spend more time evaluating unbelievable x valid arguments than unbelievable x invalid ones. This is because to
confirm validity one must consider all possible conclusions; whereas to disconfirm validity one must find only one case that does not match. Finally the third hypothesis is that arguments with believable conclusions will be solved quickly and there will be no response time difference between valid and invalid trials. This expectation follows from the premise that reasoners are not motivated to verify arguments with believable conclusions, thus they readily (and quickly) accept the conclusion regardless of the argument’s underlying structure. These three hypotheses were the foundation for Thompson et al.’s (2003) experiment.

The procedure followed by Thompson et al. was to measure the amount of time participants spent reading and evaluating a given syllogism before rendering a judgment about its validity. Secondly, after evaluating the syllogism, participants were asked to draw as many diagrams as possible that were consistent with its premises. Thompson et al. replicated the traditional belief bias effect such that believable conclusions were more likely to be accepted compared to unbelievable ones, and that this was particularly true in invalid trials where a believable conclusion often led invalid syllogisms to be mistaken as valid. In terms of response latency, the results did not support the core hypotheses because participants were observed to spend more time evaluating arguments ending with plausible conclusions than implausible ones. The key interaction Thompson et al. reported contradicted their third hypothesis in that they found that arguments with believable conclusions were solved slowly when the argument was invalid and more quickly when it was valid.

Results from the diagram task also challenged the assumptions made by the mental models and verbal reasoning theories which would claim that spending a long time on a syllogism can be equated with constructing many alternative models (as per mental models theory) or premises representing indirect knowledge and implied scenarios (as per verbal
reasoning theory). Despite documenting robust response time differences, Thompson et al. did not find that there was a difference in the number of diagrams that were created by their participants across experimental conditions.

Thompson et al.’s findings are quite at odds with the mental models theory which claims that implausible conclusions prompt re-evaluation whereas plausible conclusions lead to rapid acceptance of an argument’s validity. With some adjustment of verbal reasoning theory, these findings may however be accounted for. Thompson et al. therefore put forth the modified verbal reasoning theory which supposes that reasoners attempt to find a single model that will integrate the premises and conclusion and that the search continues until this model is found or a response deadline is attained. If the model is found before this deadline the conclusion is accepted and plausibility does not factor into evaluation. If however the deadline arrives and no model has been found the reasoner either rules that the conclusion is invalid (ignoring plausibility/implausibility of the conclusion), or they attend to plausibility and render a belief-biased judgment of validity. Thompson and colleagues further noted that the search for a valid solution is easier than for an invalid one which means that reasoners are more likely to find a solution for valid syllogisms by the deadline than they are for invalid ones. The final component to the modified verbal reasoning theory is that the response deadline is different depending on a conclusion’s plausibility – specifically reasoners are more patient with believable content but find implausible content less palatable and are therefore more rushed in their assessment. This model would therefore account for the fact that reasoners were observed to take longer processing believable than unbelievable content in invalid trials.
Selective Processing Account

Klauer, Musch and Naumer (2000) reviewed the selective scrutiny, misinterpreted necessity, mental models, and verbal reasoning accounts of belief bias. They observed that although these models hint at the issue of response bias, it is not an explicit part of these explanations. Klauer et al. described two forms of response bias and argued that this factor, acting independently of one’s beliefs, may strongly impact one’s responses on syllogistic tasks. Klauer et al. put forth a multinomial processing tree (MPT) model as a means of explaining belief bias in light of response bias.

The first form of response bias is described as an internal threshold that must be attained before a response is provided. In the case of syllogisms, this threshold would relate to the amount of evidence that a reasoner requires and the type of test that they perform in an effort to rule that a syllogism is valid or invalid. Some people may have a higher threshold than others in that they need more evidence before rendering a judgment. The threshold may also differ depending on whether one is looking to confirm or disconfirm an argument – one may need more evidence to accept as opposed to reject a conclusion, or vice versa. A key assumption that Klauer and colleagues make is that positive (confirmatory) tests are more likely performed for arguments with believable conclusions – participants are attempting to find a model in which the premises and conclusion fit so that they can accept the argument. Alternately, unbelievable conclusions are thought to promote an effort towards negative (disconfirmatory) tests such that participants attempt to construct a model where the conclusion will not fit so that they can reject the argument.

What is clear is that according to the MPT model, beliefs bias the reasoning process. In other words, beliefs about the plausibility of the conclusion determine how reasoners work
through the argument. For example, in the case of an indeterminate syllogism which could be evaluated through the construction of both confirmatory and disconfirmatory models, it is the conclusion’s plausibility that will (a) determine what the reasoner’s goal is (to confirm or disconfirm) and (b) establish the threshold or amount of evidence a reasoner will strive to accumulate before rendering a judgment about the argument’s structure. A second form of response bias is activated in situations where the threshold has not been attained and an individual lacks adequate evidence to judge a syllogism. In this instance there is a feeling of uncertainty despite a pressing need to render a judgment. A reasoner in this instance may be biased towards rendering one type of judgment (valid, invalid) over another or attempting to apply one strategy over another. For instance, facing uncertainty and needing to render a judgment may lead some individuals to rely on their beliefs about the content rather than using other external factors (e.g., atmosphere, caution, perceived base rate of valid syllogisms) to generate a response.

According to the MPT model, four processing trees can be used to represent one’s likely responses to valid x believable, valid x unbelievable, invalid x believable, and invalid x unbelievable arguments. Regardless of the argument type, it is assumed that a reasoning stage precedes the response stage. To illustrate the model we can consider the case of a valid x believable syllogism. There is a probability of $A$ that the reasoning stage will output an accurate conclusion and a probability of $1-A$ that reasoning will not yield a certain end. These probabilities are strongly related to a. the argument’s logical structure and b. the plausibility of the conclusion. If there is uncertainty and one must produce a response, there is a probability of $B$ that one will be biased by extra-logical cues (e.g., beliefs) and correctly guess that the argument is valid, and a probability of $1-B$ that one will guess that the argument is invalid.
An uncertain end to the reasoning stage makes it likely that belief bias will occur because feeling uncertain is what pushes participants into a response phase that relies on guessing and the use of extra-logical cues. Extra-logical cues include atmosphere, beliefs about the content in a proposition, as well as how likely participants think it is that an argument is valid. Klauer et al. manipulated participants’ expectations about the likely structure of arguments by telling them that the stimuli were picked from a group of syllogisms consisting of many, a moderate number, or very few valid arguments. Consequently participants came to expect that there was a high, moderate, or low probability of any given syllogism being valid. Klauer et al. proposed an equation to describe how extra-logical cues like beliefs and expectations about validity are likely to interact. Specifically, they argued that the likelihood of guessing that an argument is valid equals the product of one’s certainty that an argument is valid (high, medium, low) and the perceived plausibility of the conclusion. Thus, it is highly probable that a participant will guess “valid” when they expect that the syllogism has a good chance of being valid and when they believe the content, whereas the probability is lower when they expect that the syllogism has a good chance of being invalid and when they do not believe the content.

Klauer et al. conducted 8 experiments, some in the laboratory and others over the World Wide Web that allowed them to collect data from participants with varying degrees of logic training. They found that the base rate manipulation was effective and that it interacted with beliefs but only in the response phase. That is to say, participants always attempted to reason about the premises and conclusion and only when this process failed and they remained uncertain did they respond in a manner where their beliefs about the content and suspicions about the structure came into play. Participants who were trained in logic were more successful resolving the syllogisms through reasoning and were less reliant on guessing, however when guessing was
necessary they were no more likely to rely on beliefs or structural suspicions than novice participants. Klauer and colleagues formalized their interpretations by proposing the MPT model which falls within the mental models framework. It assumes that participants attempt to build a coherent model using the premises, but failing this that they attempt to find evidence (possibly from beliefs) to guide their guessing.

**Receiver Operating Characteristics Account**

Dube, Rotello and Heit (2011) reviewed the five contemporary models that explain deductive reasoning and account for belief bias, as discussed above. They then proposed a sixth model, the *Receiver Operating Characteristics* account, which is said to be an extension of Klauer, Musch and Naumer’s (2000) MPT model. Much like Klauer et al., Dube et al. consider the role of response bias and argue that previous research has not adequately distinguished belief bias effects that may lead to judgments of validity from response bias effects, or participants’ general predilection to endorse conclusions. A response bias effect might be controlled for by subtracting the number of false positives from the number of correct responses (hits) thereby yielding a corrected accuracy score. However, according to Dube et al. this correction may only be applied if the ratio of hits to false positives is linear when the two response types are plotted one against the other. That is to say, to apply this correction it must be true that one’s tendency to endorse a particular conclusion leads to equivalent increments in the acceptance rates for valid and invalid problems. Theoretically speaking this would yield a plot of hits against false positives where there is a constant degree of response bias, also known as an isosensitivity curve or a ROCs function. Dube et al. suggest that the threshold or amount of evidence one needs to render a valid judgment may differ between valid and invalid trials which would imply that the
ROC is actually curved and that the variances in participants’ accuracy on valid and invalid trials is non-equivalent.

In an effort to produce empirical ROCs, Dube et al. asked participants to judge the validity of syllogisms and also rate their confidence about these judgments. Based on recoded confidence ratings, Dube et al. constructed an empirical ROC contrasting hits (“valid”/valid) with false positives (“valid”/invalid) which they observed to be curved rather than linear. This led them to conclude firstly, that the belief bias effect is not a real phenomenon and is rather a Type 1 error, and secondly, that previous syllogistic reasoning accounts ought to be replaced by their new model.

Comparing the Models

Many researchers have completed studies and authored theoretical critiques that directly pit two or more of the models against each other. I will summarize these articles sequentially according to their dates of publication and highlight the key conclusion in each case in terms of which model(s) gained support (see Table 3). To facilitate our discussion in coming sections, the models will be referred to as follows: selective scrutiny (SS), misinterpreted necessity (MN), mental models (MM), verbal reasoning theory (VR), multinomial processing tree (MPT), and receiver operating characteristics (ROC).

Newstead et al. (1992) noted equal levels of belief bias on determinately invalid and valid trials. This finding accords with MN which contends that uncertainty associated with indeterminate structure is what leads to a reliance on beliefs. This finding challenges SS which predicted that belief bias would occur for determinately invalid syllogisms so long as the conclusions in said arguments were implausible and likely to promote logical reasoning. Newstead et al. (1992) went on to compare MN with MM by contrasting rates of belief bias for
single model syllogisms. According to MN, belief bias should be stronger in single model syllogisms that are indeterminate than determinate, whereas MM would predict equal levels of belief bias across single model syllogisms since only one model needs to be constructed, something that is accomplished with equivalent ease across conditions. Findings supported the latter model and were bolstered by the results of two more studies that demonstrated that the interaction between plausibility and validity reappears if multiple model syllogisms are administered, and that belief bias is attenuated by providing instructions that stress to participants the importance of applying a logical strategy. In this instance, MM is supported because Newstead et al. showed that any measure that is taken to encourage a construction of alternate models promotes greater accuracy than when a plausible conclusion discourages verification.

Table 3

<table>
<thead>
<tr>
<th>Variable Support for the Models</th>
<th>Selective Scrutiny</th>
<th>Misinterpreted necessity</th>
<th>Mental models</th>
<th>Verbal reasoning</th>
<th>MPT</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newstead et al. (1992)</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oakhill &amp; Garnham (1993)</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capon et al. (2003)</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Ball et al. (2006)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Stipple &amp; Ball (2007)</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Klauer &amp; Kellen (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
In a discussion piece published in *Cognition*, Oakhill and Garnham (1993) revisited their earlier results (Oakhill, Johnson-Laird, & Garnham, 1989) and championed MM over SS and MN based on the fact that MM explains the process of *conclusion generation* – how reasoners derive a possible conclusion from premises, whereas SS and MN describe instances of *conclusion evaluation* which is a limited account of only those cases in which a conclusion is provided to participants. Oakhill and Garnham concluded that the mental models account provides “the most convincing and parsimonious” account of participants’ responses however they also highlighted a lack of consensus in the field related to belief bias rates for multiple-model valid problems. Whereas Oakhill, Johnson-Laird, and Garnham (1989) and Newstead et al. (1992) reported that participants were no more or less accurate on believable than unbelievable valid multiple-model trials, Evans, Barston and Pollard (1983) successfully documented a belief bias effect in multiple-model problems.

Capon, Handley and Dennis (2003) set out to examine the degree to which syllogistic reasoning draws on spatial and verbal working memory resources which were assumed to derive from independent executive functioning systems. Their study was therefore an implicit comparison of VRT, MM and MPT. Verbal and spatial working memory capacities were measured by way of *simple* and *complex* word and image recall tasks, respectively. Simple tasks gauged participants’ ability to remember individual items whereas complex tasks focused on the recall of groups of items. A correlational analysis revealed that syllogistic reasoning performance was significantly correlated with participants’ performances on all of the working memory measures. The authors speculated that verbal memory resources were possibly implicated because the syllogisms were presented verbally in their first experiment. In a second experiment they compared syllogistic performance across two modes of presentation and reported that the
correlation between word span and syllogistic reasoning decreased from the verbal to the visual modes, which not surprisingly suggests that verbal working memory is integrally involved in syllogistic reasoning when the task is presented verbally.

Finally, structural equation modelling was used to test a confirmatory factor analytic model in which verbal and spatial span measures were treated as independent (orthogonal) factors, verbal and spatial span measures were combined to produce a third independent factor representing overall working memory resources, and syllogistic reasoning performance was the outcome variable being predicted. The results of this analysis showed that all three factors emerged as significant predictors, however spatial memory measures accounted for the most variance in the model followed by the verbal and general factors. Although the dominance of the spatial factor supports the mental models account, the fact that the verbal factor is also significant “undermines theoretical accounts that posit deterministic verbal or spatial processes for syllogistic reasoning” (p. 236). The authors posit that spatial and verbal working memory are both implicated because in their sample some participants predominantly use a spatial reasoning process while others rely more so on a verbal mode of reasoning.

Ball and colleagues (2006) aimed to “arbitrate” between competing models of belief bias by collecting inspection-time data derived from eye-tracking during syllogistic tasks. They assumed that longer inspection-times are associated with more challenging syllogisms and more effortful processing than shorter inspection-times. Firstly, SS would predict that participants spend substantially longer evaluating invalid than valid arguments because invalid arguments entail more processing by virtue of the fact that they more frequently lead to uncertainty (indeterminate conclusions). According to the mental models account, participants should demonstrate a prolonged *post-conclusion inspection stage* after encountering an unbelievable
conclusion compared to a believable one because it is presumably the implausible nature of the conclusion that encourages reconsideration and remodeling of the premises. Finally, the SP theory would suggest that participants will have a great deal of difficulty constructing a model when the validity and plausibility are crossed. This should therefore give rise to extended inspection times in incongruent compared to congruent trials.

Processing times did not vary between valid and invalid trials, therefore SS received no support. In terms of MM, the results challenged this model by demonstrating that in fact, the post-conclusion inspection of premises was longer in believable than unbelievable arguments. Ball and colleagues reported support for the MPT model such that participants dwelt significantly longer on incongruent trials than congruent ones. Interestingly they noted that participants spent about the same length of time perusing the premises (4 seconds) before noting the conclusion, but substantially longer reviewing the premises after noting the conclusion in incongruent trials. Ball and colleagues argued that this pattern of inspection-time data supports a backward reasoning mechanism whereby participants skim the premises initially, but use the conclusion as an anchor from which to perform logical work on the premises in a second pass.

Ball et al. (2006) provided inspection-time based evidence for backwards reasoning and the SP account when conclusions were semantic and expressed either plausible or implausible beliefs. Stupple and Ball (2007), using neutral content in a second inspection-time based paradigm, reported results that at first glance seem to challenge their earlier findings by supporting a forward reasoning mechanism. This mechanism is inherent to the MM theory which posits that premises are considered prior to the conclusion of an argument and that this order does not vary depending on the task type (evaluation versus generation). As such, a figural effect (in terms of inspection time) is anticipated because some premises entail fewer mental operations
(e.g., AB-BC) than others (e.g., BA-CB) to achieve a contiguous middle term. This figural effect is presumed to be evident no matter the task type because participants need to work with the premises to build a model that is then used to generate or accommodate a conclusion. Other theorists (e.g., Klauer et al., 2000) have challenged the MM theory by claiming that natural reasoning takes place in a backwards manner where, under ideal circumstances, a given conclusion is used as a jumping off point for either attempted validation or falsification of the argument. Only in circumstances bereft of a conclusion will reasoners be forced to rely on forward reasoning and therefore fall prey to the idiosyncrasies of syllogistic figure en route to producing a conclusion from scratch.

Stupple and Ball (2007) compared inspection-time duration for the parts of arguments in an evaluation paradigm in order to compare the assumptions associated with the forward and backward reasoning hypotheses. They reported that, in line with MM theory and the forward reasoning hypothesis, the premises of the BA-CB figure were worked over for substantially longer than those of the AB-BC figure, and that the former figure required a re-visititation of Premise 1 en route to the conclusion whereas in the latter case participants worked in a linear fashion without backtracking to Premise 1. To reiterate, these findings support the MM theory and challenge the MPT model because participants should not demonstrate variable processing times for premises according to figure if said components are deduced from the conclusion rather than integrated to build a model. Acting on the premises is what is said to activate the figural biases evidenced through inspection time measures in this study; acting on the conclusion is the assumed first step according to the backwards model, something that yields premises that need not be integrated. Additionally, the discrepancy between the Ball et al. (2006) and Stupple and Ball (2007) results suggest that forward reasoning may occur more frequently for neutral
arguments in which conclusions are not empirically relevant whereas belief-oriented conclusions may instigate a backwards reasoning style.

Klauer and Kellen (2011) critiqued the ROC model proposed by Dube et al. (2010). Dube et al. had proposed that the likelihood of response bias (tendency to affirm a conclusion) is disproportionately distributed across the conditions derived by crossing believability and validity. As such they argued that a linear relationship does not exist between hits on valid trials and false positives on invalid trials and thus that an H – F correction applied in the derivation of accuracy rates produces Type 1 errors that are mistaken for belief bias effects. To support this claim they had participants evaluate syllogistic arguments and simultaneously report confidence ratings about their judgments. They showed that different confidence ratings were associated with different tendencies to affirm the argument, a pattern that was best fit by a non-linear line. Similarly different tendencies to affirm were observed depending on the cross-over between believability and validity. The critique waged by Klauer and Kellen (2011) made one particularly important point – that binary judgments (valid or invalid) have been shown to produce linear relationships when hits are plotted against false positives; whereas the use of confidence ratings to discriminate between different levels of valid and invalid judgments (e.g., high- vs. low-confidence) may have created the nonlinear relationship.
CHAPTER 4:
DEVELOPING AN INTEGRATED APPROACH TO BELIEF BIAS

Reconciling the Models

Each account of syllogistic reasoning discussed so far has specific features and faces particular criticisms. As such it is not possible, nor is it the goal of this thesis to distinguish the best from the worst of these accounts. Some scholars have gone so far as to critique the debate that rages between proponents of different reasoning theories. Falmagne and Gonsalves (1995) expressed misgivings about the field’s affinity for pitting one theory against another. They observed that most discussions about reasoning “have advocated the exclusivity of either one or the other process in deduction, a ‘right or wrong’ rhetoric,” whereby “analyses are radicalized into extreme, opponent accounts, each claiming to explain the phenomenon in its entirety with a single theoretical language and a few constructs… and focusing the bulk of analytical and rhetorical effort on establishing the supremacy of one extreme account over the other” (p. 8). Arguing against “the questionable strategy [of] oppositional debate” waged by proponents of “universalizing theories” (p. 9), Falmagne and Gonsalves assert that “integrating descriptions of formal, meaningful, and pragmatic aspects of deductive thought,” will serve the “more compelling” goal of “capturing the deductive process theoretically in a way that is commensurate with its complexity and flexibility” (p. 8).

Thus, it is important to highlight two points that may afford for a broad reconciliation of these positions. Firstly, there is evidence to suggest that people apply different reasoning styles; therefore, more than one reasoning model may be accurate. Secondly, although each of these models offers a unique explanation of analytic processing, they each assume that this form of analysis takes place alongside a heuristically driven process of analysis. This dual-processing
assumption will be elaborated upon in coming sections, and in combination with an individual differences explanation of analytical reasoning, provides us with a foundation upon which to conclude that syllogistic reasoning may take the form of a visual and/or verbal process that is driven to some degree by our beliefs and potential response biases.

**Multiple Analytic Processes are Possible**

According to Ford (1995) it is “wrong to group all people together as though they basically reason in the same fashion” (p. 1), either by envisioning “structural analogues of the world” (Johnson-Laird, 1983, as cited by Ford, 1995, p. 3) or by engaging in a verbal process of representation. Based on a study in which participants solved syllogisms and described (verbally and/or with the use of a pen and paper) the process they had used during the task, Ford concluded that there is more than one reasoning strategy. While attempting to solve and explicate their evaluation of semantic syllogisms, some participants were observed to use “shapes such as circles or squares placed in different spatial relationships” (p. 14) while others “spoke, in various ways, of replacing one term in the premise with another” and sometimes rewrote the syllogism “as though it were a kind of equation” (p. 16). Ford reported that participants tended to demonstrate a spatial- or verbal-bias, although a minority of individuals used both strategies.

Ford’s study was criticized for having a small sample size (N=20) and for its reliance on a “think aloud” procedure which may have encouraged participants to engage in a verbal process that would otherwise have been unnatural for them. Bacon, Handley and Newstead (2003) attempted to replicate Ford’s findings after correcting for these issues, and were also curious as to whether verbal and spatial strategies would be associated with varying degrees of success on syllogistic tasks. Bacon et al. compared two syllogistic task conditions – one in which participants were asked to “talk aloud” and use a pen and paper to describe the process; and a
second condition in which they were only asked to use a pen and paper. Based on a review of the protocols produced by participants, Bacon and colleagues reported that participants demonstrated clearly distinguishable verbal and visual strategies, “verbal reasoners frequently referred to actions such as replacing, substituting, and cancelling syllogistic terms, while spatial reasoners often depicted the terms, and their inter-relationships, as groups or subsets” (p. 142). The appearance of these strategies was found to be independent of the assigned condition, thus spatial reasoners visualized more than they verbalized, even in a condition in which they were able to use both strategies. A minority of participants used both strategies at comparable rates. In terms of accuracy, no significant difference was observed between spatial and verbal reasoners when it came to total number of conclusions judged correctly. Interestingly, different error patterns were observed between groups; verbal reasoners were more likely to make mistakes when the conclusion’s form was different from that of the premises whereas spatial reasoners erred more when attempting to solve syllogisms that can be depicted by more than one diagram. Thus, although overall performance accuracy rates “present an outward appearance of universality”, a closer look at reasoners’ patterns of error reveals “such ubiquity may be superficial” (p. 157).

The predominance of one reasoning strategy over another is not just attributed to individual differences between participants; it has also been linked with qualities of the syllogism or syllogistic task. For example, Marrero and Gamez (2004) reported that argument content that inspires a reasoner to think conditionally about the causal relationship between terms inspires verbal over spatial reasoning because it is verbal reasoning, being rule-based and involving verbal substitutions of linearly equated terms, that best operates on causally related terms. The type of syllogistic task also impacts the style of reasoning. When instructed to evaluate a conclusion, participants were observed to engage in backward reasoning (emphasizing
the conclusion more than the premises) significantly more often than if asked to produce a conclusion, a task set that encourages forward reasoning where premises are surveyed at greater length (Morley, Evans, & Handley, 2004).

Neuroimaging data reported by Goel, Buchel, Frith, and Dolan (2000) and the results of a qualitative review of fMRI and patient data by Goel (2007) both show that reasoning can be spatial or verbal. Goel and colleagues (2000) scanned participants while they solved semantic and non-semantic syllogisms. Processing of semantic syllogisms was shown to be associated with unique activation in left hemisphere regions known to underlie semantic and syntactic processing. Thus, consistent with mental logic theory (Rips, 1994) and verbal reasoning theory (Polk & Newell, 1995), it seems that reasoning about syllogisms containing concrete content relies on a left-hemisphere driven verbal mechanism. Processing of non-semantic syllogisms was associated with its own unique pattern of activation in the parietal and right hemispheres. This pattern is known to underlie internal representation and spatial manipulation of information. Thus, this finding suggests that mental models (e.g., Venn diagrams, Euler circles) may be employed when participants solve abstract syllogisms, as would be hypothesized according to a mental models account of reasoning.

Goel et al. (2000) also looked at brain activation for incongruent trials in which plausibility and validity were in conflict. Behaviorally speaking, these trials were completed more slowly and with more errors than congruent trials. Goel et al. speculated that the left-hemisphere driven verbal mechanism that is sensitive to content may be the default processor and that it takes time to suppress it in order to make way for parietal and right hemisphere driven visuospatial processing. Conflict trials were found to be uniquely associated with right lateral prefrontal cortex (rLPFC) activity. Goel and colleagues speculated that this activation
may be involved with attention shifting and conflict resolution. The take home message from this study is that the left hemisphere seems to house a default, content-sensitive verbal processing mechanism that is necessary, and often sufficient, for reasoning while the parietal and right hemisphere based visuospatial mechanism is secondary, and often not called upon for reasoning. The authors concluded that the difference between the semantic and non-semantic conditions was not so much the presence versus absence of semantic content, but rather the presence vs. absence of beliefs.

Goel (2007) conducted a qualitative analysis of neural imaging data across various reasoning tasks and also assessed data derived from patients with brain lesions. He noted that there are neural systems that are sensitive to different problem-related variables. There are systems for dealing with a. familiar versus unfamiliar materials; b. conflict and belief-bias; and c. certain versus uncertain information. Consistent with Goel et al.’s (2000) findings, familiar materials (i.e., semantic propositions) recruited language processing regions in the left frontal and temporal lobes which are known to be implicated with situation-specific heuristics. Unfamiliar materials (i.e., non-semantic propositions), on the other hand, recruited visuospatial processing regions bilaterally in the parietal lobes. These regions have been previously implicated with rule-based applications via a more deliberate and slow processor. Conflict and belief-bias were shown to implicate rLPFC regions when such conflicts were detected and resolved accurately and the ventromedial prefrontal cortex when responses were belief-based and non-logical. With specific reference to patient data, lesions in the left hemisphere (temporal or frontal) were shown to be reliably linked with diminished reasoning, and lesions in the rLPFC were associated with failures to detect conflicts or to deal with inadequate information carefully. Goel noted that these data ought to move scholars away from “the sterility of the mental models
versus mental logic debate,” (p. 441) given that both types of processing are noted during reasoning tasks. Prado et al.’s quantitative meta-analysis (2011) reached similar conclusions.

**Analytic Processing is Distinct from Heuristic Processing**

Although there is substantial debate about the specifics of the reasoning process, the six models share in common the assumption that the analytic processing system is separate from a second, heuristic system. “Contemporary theories of belief bias are couched within a dual-process framework” (Stupple & Ball, 2008, p. 169) which assumes that System 1 (heuristic processor) acts rapidly and unconsciously, while System 2 (analytic processor), tapping working memory, more slowly mulls over problem details and applies rules en route to rendering a judgment (Evans & Curtis-Holmes, 2005). As Marrero and Gomez (2004) noted, heuristic processing is superficial processing whereby “people base their answers on the fast and frugal heuristic performing on the surface features of the problems” (p. 168), which in the context of syllogistic reasoning eclipses System 2’s proclivity to “reason by representing the information” (p. 168). Syllogisms that present empirically plausible or implausible content are the domain in which Systems 1 and 2 collide; beliefs about the world essentially ‘tip off’ the heuristic system which can lead to within-participant conflict (Evans et al., 1983) in cases where realism does not match with logical structure.

**The Heuristics System**

The heuristics system is said to act on an argument’s superficial structure and content. In so doing it automatically generates a judgment about the argument’s validity that is based on its surface qualities, for example the atmosphere invoked by its premises or the empirical value of its conclusion. Atmosphere and the plausibility=truth principle that underlies belief bias are but two examples of factors that may drive the heuristic system.
Tversky and Kahneman (1974) described representativeness- and availability-driven shortcuts that are active when people feel uncertain about everyday events and that lead to an abandonment of deliberate and rule-based reasoning en route to an ultimate judgment. Although these heuristics were not directly applied to syllogistic reasoning by Tversky and Kahneman, they are likely to shape syllogistic reasoning by making some conclusions seem more plausible or probable than others, a feeling that could translate into one’s ultimate evaluation of the argument’s validity. Based on the representativeness heuristic, participants deem statements that contain highly similar components plausible. Thus, even if two arguments are valid, the one whose conclusion alludes to similar (e.g., creative people are famous musicians) rather than less commonly related categories (e.g., creative people are successful accountants) will be more likely deemed to be valid.

The availability heuristic encourages people to judge statements based on how easy it is to generate examples of specific instances. Related to syllogistic reasoning, this heuristic could also lead to two valid conclusions being judged differently. For instance “elderly and inactive people are heart attack victims” is likely to be judged as valid more often than the statement “young and active people are heart attack victims,” because it is presumably easier to recall older, inactive individuals who have suffered heart attacks than their younger, active counterparts. Memory recall factors heavily into this style of judgment so some propositions are more prone to this than others, depending on how easy it is to remember specific instances. For instance, if a proposition is emotional then it may be easier for an example to be recalled because of salience, and if the memory is recalled quickly and with relative ease, this feeling itself may boost a participant’s likelihood of concluding that an argument is valid.
 Whereas Tversky and Kahneman’s (1974) heuristics are content-driven, Chater and Oaksford (1999) described heuristics that stem from structural features of the syllogism. They contended that participants apply probability rules to the surface qualities of syllogisms, thereby wielding “the calculus of uncertain reasoning” as opposed to engaging in “logical deduction, the calculus of certain reasoning” (p. 195). Three heuristics for generating a possible conclusion, and two others for testing the conclusion’s likely accuracy were outlined by the authors. According to this model, the heuristic processing system is often exclusively engaged, being used “in place of complex and sometimes computationally intractable optimal [logical] strategies, which are therefore denied any cognitive role” (p. 207).

According to Chater and Oaksford, the heuristics used to generate a conclusion each rely on the relative informativeness of premises in syllogistic arguments. The min-heuristic leads to the selection of a conclusion with a quantifier (all, some, etc.) that matches the quantifier of the least informative of the premises (akin to Wetherick & Gilhooly’s matching principle). The p-entailer heuristic also prompts the selection of a conclusion’s quantifier; in this case what is chosen is a quantifier that is necessitated (entailed) by the conclusion specified by the min-heuristic. That is to say that in the case where the min-heuristic leads to a preference for a conclusion quantifier of the form All X are Y, in the absence of this conclusion quantifier, the next most preferred quantifier will be one that is entailed by the statement All X are Y – specifically, that Some X must be Y. The final, attachment-heuristic leads individuals to decide on the order of terms in the conclusion, an order that suggests that the conclusion’s end-term ought to be the subject addressed in the min-premise. Once a conclusion is generated two further heuristics are used to test how likely it is that said conclusion is accurate. Rather than following an analytical route – for instance attempting to envision mental models that would falsify or
validate a given conclusion, Chater and Oaksford assume that a max-heuristic inspires trust about a given conclusion in proportion to how informative its most informative premise was. This is therefore a heuristic that biases the individual to assuming that a syllogism with informative premises is likely to have a valid conclusion. The O-heuristic encourages individuals to distrust uninformative conclusions of the form Some A are not B (a prediction that discords with the matching principle which would not discriminate between equally conservative “Some are” and “Some are not” statements).

Also describing a structurally-based heuristic, Wetherick and Gilhooly (1995) argued that the matching principle also contributes to heuristic processing whereby participants select a conclusion that matches the logical form of the more conservative of two premises. A conservative premise is said to be one that makes a claim about the least number of people, therefore a statement of the form “some A are B” is more conservative than the liberal statement that “all A are B” and is consequently more likely to be used as a heuristic cue.

In each of these cases, participants attempt to follow a rule, albeit one that is non-logical. As Kahneman (2011) expressed, System 1 tends towards over-simplification whereas System 2 system houses our logical toolbox. We might apply logical tools to difficult questions if it weren’t for the hasty actions of System 1 which often provides “off-the-shelf answers” (p. 99) to easier questions than the ones posed by the task at hand. Together, the sections above have catalogued the types of answers stocked on the heuristic shelf.

**Evidence for Dual-Process Systems**

Goel and Dolan (2003) used fMRI and provided evidence for the existence of distinct heuristic and analytic systems. They wanted to determine how syllogistic reasoning about
congruent, incongruent, and neutral arguments pertains to brain activation. A comparison of correct response trials showed that:

i. the left temporal lobe was more active in trials where beliefs facilitated responding than in those where beliefs were uninvolved, a finding that corresponds with previous work showing that this brain area is linked with retrieval of semantic information (e.g., Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997, as cited in Goel & Dolan, 2003),

   ii. activation in the bilateral parietal lobe was greater for trials where logical reasoning dominated (neutral trials) than in trials driven by beliefs. Again, this is a finding that corresponds with previous work; the parietal system is known to be active when efforts are underway to internally represent and manipulate spatial information (e.g., Laeng, 1994, as cited in Goel & Dolan, 2003).

   iii. correct responses on incongruent trials in which inhibition of one’s beliefs is imperative for accuracy were accompanied by rLPFC activity. This area is known for its role in cognitive monitoring – both in terms of detecting and resolving conflicts between one’s intentions and one’s circumstances (e.g., Goel et al., 2000; Fink et al., 1999, as cited in Goel & Dolan, 2003).

   iv. incorrect responses on incongruent trials in which beliefs have seemingly triumphed over logical reasoning corresponded with activation in the ventral medial prefrontal cortex which is an area that seems to be connected with affective processing and the limbic system (e.g., Goel & Dolan, 2001, as cited in Goel & Dolan, 2003).

In sum, Goel and Dolan’s results support the view that there are two separate systems underlying heuristic and belief-oriented versus analytic and logic-oriented processing. In addition, they showed that in conflict situations where either system could be employed to solve
a problem, logical processing is more likely if the conflict is explicitly recognized and one’s intentions to behave logically are reinforced via cognitive monitoring whereas heuristic processing is more likely if the affective system kicks in, possibly because problem content strikes an emotional chord with the individual.

Evans and Curtis-Holmes (2005) also provided evidence for a dual-process theory of reasoning by highlighting the existence of the heuristic system. They accomplished this by introducing a time limit to syllogistic reasoning trials which ostensibly reined in the slower-acting analytic system. Participants were shown to respond differently to syllogistic tasks depending on the length of time they had available for problem-solving. Free-time trials were associated with substantially higher accuracy rates and a lower frequency of belief bias than time limited trials. The most important difference that Evans and Curtis-Holmes described was that for congruent trials there was little difference in accuracy between the free-time and time-limited trials whereas a time limit was particularly detrimental for incongruent trials in which an analytic strategy is imperative for accuracy and content is clearly a misleading cue. Interpreting this difference, Evans and Curtis-Holmes suggested that the time limit may lead to more errors because it impedes a logical search for alternative models and/or necessitates a reliance on surface cues (content) because there is inadequate time for an in depth consideration of the problem’s structure.

Reverberi, Rusconi, Paulesu and Cherubini (2009) also demonstrated that task demands differentially recruit the analytic and heuristic systems - time constraints led to greater reliance on heuristic strategies than analytic ones. These authors extended Evans and Curtis-Holmes’ main finding by distinguishing between structurally-driven and content-driven heuristic strategies. Specifically, when participants had less than 20 seconds to evaluate thematic
syllogisms, atmosphere was shown to be the primary surface feature that was assessed heuristically, whereas beliefs about content were a secondary heuristic cue that mostly factored into judgments of incongruent syllogisms. Reverberi et al. were able to demonstrate this graded effect of heuristics by designing a study in which the two heuristic strategies were pitted against each other. Participants’ performance on atmosphere-belief conflict trials, in which beliefs prompted one form of responding and atmosphere prompted an opposite response, indicated that priority was afforded to a structurally-oriented rather than a knowledge-driven heuristic approach. An interesting difference that was noted about the two types of heuristic strategies was that the structural heuristic disappeared when participants were given more time to complete the task, whereas beliefs were likely to bias responding under limited and unlimited time conditions.

Tsujii and Watanabe (2009) provided evidence for dual process systems by designing a condition that would impede System 2 but not System 1. They applied a dual task paradigm to the study of belief bias based on the assumption that System 2 relies on working memory whereas System 1, being an automatic and intuitive processor, does not have this reliance. Near-infrared spectroscopy (NIRS) was employed as a means to examine the neural correlates of task performance; this imaging technique assesses cortical hemodynamic responses by tracking the ease of passage of near-infrared light through tissue. Oxygenated and deoxygenated hemoglobin have different absorption spectra, thus it is possible to infer the concentration level of oxygenated hemoglobin which is itself an index of cortical activation (Obrig et al., 2002, as cited in Tsujii & Watanabe, 2009). Performance accuracy and cortical activation were compared during congruent and incongruent trials under high- (monitor the location of a shape relative to two previous trials) and low-load (monitor the location of a shape on current trial only). Behaviorally speaking, accuracy was worse on high- compared to low-load trials and this
diminishment in performance was particularly marked for incongruent trials. Given this effect, Tsujii and Watanabe have added to the evidence suggesting that System 1 and 2 are independent.

In terms of neural activity associated with such conflict trials, Tsujii and Watanabe reported that in participants who managed to deduce an accurate judgment, presumably having inhibited beliefs and gone with logic, there was a significant positive correlation between accuracy and activation in the rLPFC. When such conflict trials were encountered under high-load conditions, rLPFC activation decreased for all participants; however, in participants who still managed to produce an accurate judgment, rLPFC activity was markedly higher than in participants who failed to do so. The authors concluded that rLPFC activity acts as a mediator during belief-logic conflict, specifically inhibiting the heuristic system so that the analytic system can go ahead and process the argument. This interpretation accords with previous research highlighting the rLPFC’s potential role as an inhibitor (e.g., Aron et al., 2004, as cited in Tsujii & Watanabe).

The above authors established that System 1 and 2 seem to be distinct processors. Additionally, Goel and Dolan (2003) established that System 1 and System 2 carry out their processing by harnessing separate neural zones, the left temporal or bilateral parietal lobe, respectively, and that each system has its own sponsor or accelerator, the ventral medial or right lateral prefrontal cortex, respectively. In terms of a braking system, Tsujii and Watanabe (2009) also implicated the rLPFC as being a possible inhibitor of System 1. Studies in the coming section work from these foundations and delve into the question of how System 1 and 2 interact.

Heuristic and Analytic System Interaction

A variety of explanations have been offered about the nature of the interaction between Systems 1 and 2. While some explanations have identified System 1 as the primary processor,
others have cast Systems 1 and 2 as equally important processors that handle discrete types of information. In contrast, Handley and colleagues (2011) proposed that Systems 1 and 2 greatly overlap and cannot be distinguished based on what they process but rather how they act upon this common information. Finally, an individual differences perspective (e.g., Stupple et al., 2011) contends that some people are System 2 specialists whereas others are not.

*System 1 predominates.* According to results reported by Shynkaruk and Thompson (2006), **System 1 seems to be the dominant system by default.** Heuristic processing is fast, it is easy and left to our own devices, System 1 activation may eclipse the use of logical analysis, or even harness System 2’s ability to rationalize in an effort to back-up illogical, biased judgments. Shynkaruk and Thompson (2006) compared accuracy and confidence ratings about syllogistic judgments. They showed that rather than reflecting participant’s analytical performance, confidence ratings were generated heuristically and relied on the plausibility of the arguments in question. Implausible content inspired low confidence which led participants to change their answers when given the opportunity to re-evaluate syllogisms. Answer-changing, when spurred on by low confidence, tended to produce a higher error rate compared to that associated with an initial set of responses. Plausible content on the other hand inspired elevated confidence, and participants were not observed to change their answers when given the opportunity to re-evaluate syllogisms. Instead they reported greater confidence than they initially had. These findings suggest that the heuristic system impacts the analytic system, and not the other way around. In this study uncertainty or certainty about surface features led to the respective abandonment of or overconfidence in one’s syllogistic judgment.

Shynkaruk and Thompson (2006) went so far as to claim that System 1 is dominant to the point that intuitive trust of syllogistic content engages System 2 in a rationalizing process
whereby an initial, biased judgment is justified again and again, and comes to be trusted to an even greater degree. While “people mistakenly assume that their thinking is done by their head [System 2],” Shynkaruk and Thompson demonstrated that “it is actually done by the heart [System 1] which first dictate[d] the conclusion, [and] then command[ed] the head [System 2] to provide the reasoning that [would] defend it” (Anthony de Mello, 1931-1987). This characterization of Systems 1 and 2 is in kind with Kahneman’s view that System 2 is inherently lazy. Accordingly, belief bias happens because System 1 automatically generates suggestions that are forwarded to System 2, which by nature wants to endorse these impressions and feelings.

A review of work by Morsanyi and Handley (2012) suggests that System 1 may handle some of the responsibilities shouldered by System 2. Thus, because System 1 has its own discrete function (belief-processing) and dips its hand into System 2’s pie (rendering intuitive conclusions about logical structure), it is the dominant system. Morsanyi and Handley (2012) provided evidence for the intuitive detection of logical validity which extends the position that System 1 may dominate System 2 such that logical judgment may not always happen due to conscious, effortful processing and may at times be the result of unconscious, intuitive processing. The authors assessed participants’ intuitive evaluations of syllogisms by asking them to rate their “liking” of conclusions. In line with the processing fluency hypothesis (e.g., Winkielman & Cacioppo, 2001, as cited by Morsanyi & Handley), it was assumed that “liking” ratings tap intuitive estimations of coherence and that positive affect is experienced when an unconsciously constructed model is able to accommodate a given conclusion. This emotion was thought to be experienced consciously, yet to be free floating in that individuals do not know its implicit source.
In their first experiment Morsanyi and Handley showed that participants liked valid conclusions more than invalid conclusions even when they had been explicitly instructed not to use a logical strategy to solve the syllogisms and therefore failed to review the argument’s premises. This suggests an unconscious, intuitive sensitivity to logical structure that was active when System 2 was deliberately disengaged. In a second experiment the authors showed that participants liked valid conclusions more than invalid conclusions, even when these valid arguments had difficult, working memory tapping figures that resulted in less accurate logical evaluations. Figural difficulty was a problem for System 2 but did not dissuade heuristic liking, which suggests that different forms of processing are entailed for logical validity judgments compared to heuristic estimations of logical structure. A third study showed that “liking” was linked with a preference for valid over invalid arguments only when felt emotion was free-floating. In the cases where situational factors (music playing in the background) existed to which positive emotions could be misattributed, participants’ “liking” of valid syllogisms tended to compare to their rate of “liking” for invalid syllogisms. It is as though they no longer used affect as a heuristic cue because they no longer associated their positive feelings with the argument. In a final study the authors introduced subliminal affective primes which were shown to increase “liking” of conclusions when they were positive and decrease “liking” when they were negative. This supports the view that “liking” is sensitive to affective state, something that is normally derived during a syllogistic task from the unconscious processing of fluent or influent problems, but that can also be induced artificially through priming.

Finally, results of a study completed by Dijksterhuis, Bos, Nordgren and van Baaren (2006) support the argument that System 1 is dominant over System 2, at least in the case of complex tasks. Upon comparing unconscious (System 1 driven) and conscious (System 2 driven)
deliberation leading to a decision, this Dutch team of researchers showed firstly, that participants who made ‘gut based decisions’ were more successful at identifying a good from a poor vehicle when they had to consider multiple features of that vehicle. Secondly, in terms of attitudes towards products, Dijksterhuis et al. found evidence that people derive more satisfaction from simple products they have purchased following conscious deliberation and complex products they have purchased spontaneously based on gut instincts. This study provides some support for the view that System 2 has its limits and that System 1 is better able to guide judgments when there are many features and task demands are therefore elevated.

_Systems 1 and 2 are processing partners._ Some dual-processing perspectives emphasize a belief-first style of processing and suggest that System 1 is dominant and invulnerable to System 2. In contrast, other perspectives are based on the view that Systems 1 and 2 run in parallel and impact each other. Systems 1 and 2 are described as running in tandem by Stupple and Ball (2008) who argued that although intuitive judgments may be easy to arrive at, they are not automatically applied in a problem situation and instead are weighed against analytic judgments. Thus, it is in the absence of analytic conclusions that System 1’s intuition is most likely to be applied. This perspective stems from the observation that participants were worse at solving incongruent syllogisms with difficult figures than easy ones, a finding reported by Stupple and Ball (2008) who described this condition as one in which problem complexity tips the balance so that heuristic processing furnishes the answer because it is so challenging to produce one by way of analytic processing. Figural complexity is essentially a roadblock that impedes System 2 but not System 1, yet as Stupple and Ball noticed, conflict trials were associated with the longest dwell times which was taken as evidence for a parallel process where System 1 quickly produces a belief-based response, yet individuals wait to hear from System 2 before rendering a verdict.
The idea that System 1 runs independently and at times to the exclusion of System 2 was similarly refuted by Stupple and Waterhouse (2009), who designed a set of syllogisms containing double-negatives. These problems are known to be more challenging to interpret than their positive counterparts, yet they are comparable in terms of logical structure. Stupple and Waterhouse expected that the double-negation would be a road block for the analytical system but not the heuristic system because it would necessitate the application of an additional rule in order to reverse the negation, a task that harnesses System 2. Thus, it was hypothesized that reasoners would spend longer and perform less accurately on trials containing double-negation. The results indicated that System 1 and 2 must have been working in tandem. The fact that more valid than invalid conclusions were accepted, and that negated as well as invalid problems were pondered at length compared to their respectively positive or valid counterparts, was taken as evidence for System 2 activation. Heuristic processing was evidenced by the fact that conclusions were more likely to be accepted when their form (negated or not) matched that of their premises. Finally, evidence for the interaction between Systems 1 and 2 was demonstrated by the fact that negated problems took longer to evaluate when their conclusion and premises were of opposite types (e.g., negated vs. positive). This suggests that System 1 assessed the problem for fluency, applying a matching heuristic that then dictated the amount of elaboration required by System 2.

An article by Sperber et al. (2010) also implied that System 1 scans information and under certain conditions hands it over to System 2 for processing. What can be taken from their article is the idea that System 1 seems to affirm incoming information by default, and yet simultaneously scan for self-relevance and signs of deceit that would then inspire more vigilant and elaborated consideration of information, presumably at the hands of System 2. This
confirmation bias was documented by Gilbert, Krull and Malone (1990) by way of an experiment in which participants processed a new piece of information completely or were interrupted. Interrupted processing led to a higher rate of false positives when participants were later asked to evaluate the veracity of inaccurate information – as though they extended “interpretive charity” (Sperber et al., p. 363) when asked to judge inadequately reviewed material. No comparable increase in affirmation for accurate information was observed. This suggested to Gilbert et al. that interruptions prevent participants from moving away from their default treatment of content towards the effortful application of a ‘fiction tag’ to implausible content. Additionally, Gilbert and colleagues demonstrated that if participants were informed that material was likely to be implausible, they were still likely to affirm the content if their processing trial was interrupted. This suggests that confirmation bias is an automatic heuristic and that if time is limited the more effortful process of ‘unbelieving inaccurate content,’ presumably a System 2 driven tagging process, is not given time to run its course.

Hasson, Simmons and Todorov (2005) qualified Gilbert et al.’s description of confirmation bias by demonstrating that this tendency exists for neutral, uninformative content, but disappears in the face of content that is personally relevant by virtue of its informative nature and pertinence to one’s belief systems. Hasson et al. reported that interrupted processing led participants to frequently affirm uninformative content that they had been told was false (e.g., person A drinks tea for breakfast) but that they had no problem denying informative content that they had been told was false (e.g., person A is a liberal). In terms of the conditions, saying a person is a liberal is said to be self-relevant because one is likely to have an opinion about politics whereas saying a person drinks tea is said to be non-informative and generally neutral since most people would not have an opinion about morning tea drinking. In terms of findings,
what is key here is that the default confirmation bias was shown to be active for uninformative false content but absent in the face of false content pertaining to one’s beliefs. In this latter case it seems that belief suspension was initiated and completed despite the interruption. Given that participants did not have a lot of time to review the content before being interrupted and then asked to evaluate its veracity; it seems unlikely that belief suspension was the product of System 2 activity. Rather, what is implied by Hasson et al.’s findings is that although belief suspension is conventionally seen as a System 2 process whereby extended grounds for ‘unbelieving inaccurate content’ are constructed, this process may at other times be heuristic in that “people may spontaneously form a general impression instead of making an elaborative inference” (Hasson, et al., 2005, p. 570).

*System 2 can act on System 1.* Sloman’s parallel process model (1996) accords with the view that Systems 1 and 2 are processing partners by highlighting System 2’s ability to inhibit System 1. In particular this perspective supposes that both systems attempt to generate a response and that although System 1 “always has its opinion heard and can defuse a rule-based response,” System 2 in its own right, “has some capacity to suppress [System 1]” (as cited in Stipple & Ball, 2008, p. 172). Elaborating on Goel and Dolan’s (2003) and Tsujii and Watanabe’s (2009) work implicating the rLPFC in System 1 inhibition, a number of studies provide evidence that System 2 can override System 1.

For one thing, it seems that belief bias and the application of heuristic strategies can be attenuated by the mere act of providing instructions to participants that emphasize the importance of exercising a logical strategy. Dickstein (1975) showed that instructions diminished the conversion effect, one product of heuristic processing. Evans, Newstead, Allen and Pollard (1994) focused on the belief bias effect and highlighted that only a particular type of instruction
was effective for *reducing*, but not entirely *cancelling out* heuristic processing in the form of belief bias. A type of verbal instruction that was not found to improve the rate of logical responding focused on logical necessity; despite being verbally instructed to only accept conclusions if they *must* be true given the premises, participants’ responses continued to demonstrate a robust belief bias effect whereby many invalid arguments were accepted as true when embedded with believable content. A broader set of instructions, explaining the logical meaning of “some” (implies at least one and possibly all) and emphasizing the importance of basing one’s judgment only on given information, was successful in that it seemed to induce a drop in acceptance of invalid-believable conclusions. The fact that belief bias was reduced but not eliminated by this second type of instruction led the authors to conclude that it is very difficult for people to suspend System 1’s belief-oriented processing but that people have some degree of conscious control over reasoning and given adequate motivation (instructions stressing the importance of syllogistic structure and the meaning of the “some” qualifier), will engage in an effortful search for counter-examples. Evans et al. suggest that providing instructions leads participants to suspend their everyday style of reasoning which relies on probabilistic evaluations of content, and to engage in a less natural form of analysis – syllogistic reasoning.

De Neys and Franssens (2009) showed that belief inhibition is a key part of conflict resolution when syllogistic content is incongruent with logical structure. Additionally, they demonstrated that successful belief inhibition translates into content that is blocked and rendered temporarily inaccessible. Finally, they shed light on the question of why inhibition failure occurs. Inhibition failure occurs not because the process of inhibition was uninitiated (as would occur if one felt their beliefs were ‘safe’ and unbiased), but rather because the initiated inhibition process was not completed. This finding recalls one of the key findings from Gilbert et al.’s (2000) study
which was that participants tended to confirm content, a default towards ‘aye-saying’ that could not be overridden by participants if they were interrupted during content processing. This implies that belief inhibition is more than a mere questioning of content. Rather, it involves blocking content, or as Gilbert et al described, effortfully applying a ‘fiction tag’ to content in order to highlight that the content in question is extraordinary and not to be confirmed by default. Participants in De Neys and Franssens’ study completed a syllogistic task followed by a lexical decision task. Those who accurately solved incongruent syllogisms, presumably by inhibiting their beliefs, were slower at deciphering words with semantic relation to the syllogistic content, than those who failed on the incongruent trials. This indicated that belief inhibition accompanied System 2 processing, and secondly, that inhibition led to content blocking. Interestingly, success or failure on the conflict trials led to comparable performance for the detection of neutral words, something that indicates that one’s general ability to access information is unimpaired, and rather that it is a very specific deficit that follows successful belief inhibition.

Finally, a comparison of the best (group average of 93%) and worst (solved 0%) conflict solvers revealed that the above pattern was consistent between groups which lends support to the view that one initiates the inhibition process when there is a logic-belief conflict, but that sometimes this effort fails. As De Neys and Franssens’ title implies, the process of belief inhibition does not always win but always plays a part. If one were not to initiate belief inhibition, as might be expected in participants with such dismal success rates, then subsequent lexical decision performance ought to have been unimpeded for words related to the primed syllogistic content.

Systems 1 and 2 process the same information, albeit in discrete ways. So far we have looked at accounts of reasoning that assume System 1 is fast and content-oriented while System
2 is slow and structurally-oriented. As has been noted, the main debate has been about whether System 1 or 2 is dominant and how content and structural processing may proceed in the face of problems that pit the two systems against each other. Alternative perspectives have explored the idea that Systems 1 and 2 do not distinguish between content and structure and instead process both forms of information.

In the 1990s Falmagne and Gonsalves proposed that an integrated approach to content and form would provide the best explanation for reasoning. According to this perspective Systems 1 and 2 should not be distinguished based on what but rather how they act. Falmagne and Gonsalves’ perspective took direction from the ideas that content factors into reasoning rather than being omitted, and that reasoning is belief updating (Sperber & Wilson, 1986; Harman, 1986, as cited by Falmagne & Gonsalves, 1995). Sperber and Wilson treated content as material upon which a deductive device operates, rather than material that distracts or jams said processor. Specifically, “pragmatic assumptions or world knowledge” are thought to “enrich a problem’s representation by adding additional premises that are then processed by a deductive component” (p. 14). In this way the deductive device acts on rich, contextualized material; it does not factor out content or engage in a solely abstract process. Falmagne and Gonsalves couched their proposal that syllogistic reasoning is not independent from belief-oriented judgments in Harman’s (1986) theory that reasoning is belief revision whereby the key goal is to achieve a coherent world view. Reasoning ought to be situated “within the natural course of belief updating” (p. 8) because it “integrates the various elements of a person’s knowledge base rather than giving precedence to abstract, formal knowledge” (p. 8). Continuing along this vein, beliefs are said to be attitudes towards propositions, they are “constitutive of a rational activity” (p. 17); they are not the base for the non-logical and heuristic production of distortions.
Handley, Newstead and Trippas (2011) provided empirical backing for the view that belief-based, associative processing is not the exclusive responsibility of System 1, and that structurally-oriented, rule-driven logical processing is not restricted to System 2. According to their account of reasoning, System 1 carries out fast and intuitive processing that is sensitive to both content and structure, whereas System 2 is a deliberate, effortful processor that handles more in depth considerations of both content and structure. These conclusions were grounded on observations about how accurately and how quickly participants were when judging the plausibility or logical structure of incongruent (belief-logic conflicting) syllogisms. In terms of processing times, participants took longer to judge the content than the logical structure of incongruent syllogisms which implies that time-consuming content analysis may be conducted by System 2. Secondly, participants were observed to make more errors when judging the content than the logical structure of incongruent syllogisms. This implicates System 1 as an implicit logical processor because given that participants were instructed to ignore structure and focus on content (thereby de-activating System 2), System 1 must have been intuitively assessing the logical nature of incongruent syllogisms or the conflict trials would not have differentially impacted performance relative to non-conflict trials.

System 2 specialists are a distinct ‘type’ of person. Stipple et al. (2011) worked from the premise that there is a subset of reasoners who are highly motivated to avoid the “fundamental computational bias” (Stanovich, 1999, as cited in Stipple et al., 2011, p. 933) of rendering a verdict about syllogistic structure that was coloured by its content. Sá, West and Stanovich (1999) suggested that these reasoners have high cognitive ability, are skilled at decontextualisation, and tend to think actively and open-mindedly relative to individuals who fall prey to belief bias. The authors hypothesized that this subset of reasoners would be
distinguishable from other reasoners because they would be highly accurate when evaluating syllogisms, and would take longer than average to evaluate syllogisms of the believable-invalid type. In essence, these reasoners are thought to be highly skeptical and are expected to expend considerable effort to ensure that they are not biased by content, even when that content is reassuringly plausible. Upon comparing ‘high’ and ‘low’ logic groups, the authors demonstrated that the ‘high’ logic group tended to take longer when solving all syllogisms than those in the ‘low’ logic group. This effect was not interpreted as a difference in ‘cautiousness’ because there was a variation in response times between syllogistic trial types with the ‘high’ logic group taking substantially longer on believable-invalid trials relative to the other trial types. The authors concluded that the ‘low logic’ group were engaging in heuristic processing driven by syllogistic content whereas the highly accurate group conducted an assiduous analytic assessment aimed at uncovering the structure of the argument.

Recent evolutionary perspectives on reasoning also emphasize that System 2 is subject to individual differences (Evans, 2006). Interestingly, early applications of evolutionary theory to reasoning such as Rips’ (1994) mental logic hypothesis focused nearly exclusively on System 1 which was presumed to house evolutionarily old tools that do not vary between individuals (as cited in Cosmides & Tooby, 1996). Rips took inspiration from Chomsky’s explanation of grammar by positing that subjects can solve infinite sets of syllogisms embedded with various contents because they have an innate system for processing such structures; “all normal members of the human species are equipped with an innate and extensive set of deduction principles that govern how we reason” (as cited in Cosmides & Tooby, 1996, p. 448). Thus, individual differences do not exist because everyone is essentially programmed in the same ways. A shift from presumed universality towards an emphasis on individual differences was inspired by
several observations that challenged Rips’ hypothesis. For one thing, participants often commit errors when solving even simple deductive problems and demonstrate widespread irrationalities (including belief bias). One would assume that if a logical toolbox were innate and accessible that it would always be recruited and ensure a consistently high level of performance regardless of the task parameters. Rips attempted to account for this problem by proposing that one’s innate set of logical tools are doled out at random from one task to another, and that some tools are highly accessible whereas others that are more sophisticated are only available 20% of the time.

Cosmides and Tooby (1996) critiqued Rips’ explanation, noting that it seems unlikely that a system aimed at generating adaptive behavior would vary so radically from task to task, and secondly, that performance from trial to trial on reasoning tasks does not show the degree of variation that would be expected if Rips’ explanation were accurate. In fact, some task types are performed with high reliability which should not be the case if logical tools are applied at random. Cosmides and Tooby did not dismiss the evolutionary perspective and instead contributed the idea that certain content types are evolutionarily significant and inspire specific types of reasoning. Thus, Cosmides and Tooby concluded that reasoning includes some innate logical procedures, and also a battery of specialized inference procedures that are sensitive to the content of problems and that show greater variance between individuals.

Stanovich and West (2000) contended that System 1 inspires universal responses that are normatively rational because they support the innate goal of self-preservation whereas System 2 permits specialized responses that support personal utility and consequently that differ between individuals. An example may help to elucidate Stanovich and West’s evolutionarily based explanation of dual-processing systems. System 1 pushes individuals to respond in kind by applying the same set of heuristics to situations that they encounter. This means that the same
response is likely to result in a given problem situation for most individuals. In the case of a syllogism, the same validity judgment is likely to be heuristically generated if System 1 is guiding the evaluative process. No differently, it would be System 1 that leads most antelopes to heuristically assess the alarming approach of a predator and respond in kind by fleeing en masse. System 1 inspires universal reactions that support basic goals. Not standing out keeps one safe, yet it limits one’s possibility for individual reward; something that is optimized instead by System 2. System 2 can produce the anomalous response which boosts personal potential and may lead to individual reward. For instance, the reasoner who can de-contextualize is the one who will be rewarded in a society that values syllogistic reasoning ability.

To round out our discussion it is important to highlight a key issue raised by Evans (2006) that challenges the application of evolutionary theory to explain reasoning and individual differences. In a nutshell, the evolutionary perspective presumes that the sophistication of System 2 varies between individuals and therefore that some people are more successful reasoners than others. This view is supported by findings that show that general intelligence, age, and executive control relate to syllogistic reasoning ability and vulnerability to belief bias. Yet, as Evans points out, just because an individual with higher intelligence performs better on syllogistic tasks than someone of lower intelligence does not provide conclusive evidence that their reasoning infrastructure is superior or more evolved. Rather, it could be that reasoning infrastructures are on par with each other and the one with higher intelligence taps these resources more often than those of lesser intelligence.
CHAPTER 5: THE CURRENT PROJECT

A review of the literature has revealed that syllogistic reasoning has been discussed and debated for centuries. In an effort to explain why people commit syllogistic reasoning errors, classical scholars have tended to focus on structural aspects of the syllogism and have under-emphasized the role of syllogistic content. A reassessment of the relative importance of structure and content is warranted given that the spirit of times past, be it the British Enlightenment’s valuing of rationality or New Criticism’s prioritizing of structure, have to a large degree dictated how psychologists have defined and studied these variables. An attempted rapprochement of structure and content has led to an elaborated description of the conditions that inspire belief bias as well as a comparison of different mechanistic accounts of this effect. Although a debate continues between proponents of these models, my attempts to reconcile them has led to an emphasis on their shared qualities – specifically, that research has demonstrated that multiple analytic processes are possible and secondly, that each of the accounts can be couched in a broader, dual-process model. A final piece of background related to this dual-process model has been developed; the heuristics system was described, evidence for dual-process systems was presented, and a variety of possible forms of interaction between Systems 1 and 2 were discussed.

These chapters have provided a foundation for the current project in which I assume that content and structural variables contribute to reasoning. I am aiming to understand how belief bias is influenced by the interaction between these two variables and intend to explain this interaction by way of a dual-process perspective. My project is not a systematic comparison of the belief bias models; rather I am assuming that more than one model may be required to account for this effect. In fact, my key findings suggest that no single model is accurate.
There are two final streams of research that must be described as they are particularly relevant to my dissertation. Their convergence represents the merging of theory and method. The theoretical stream of research that I will describe below formalizes the integrative approach that I am taking. It assumes that content and structure interact and that both variables can help or hinder reasoning. Thus, I would contend that when content acts on reasoning, it does not always bring about a deleterious end to the logical process. Certain types of content, experienced under certain types of conditions, may help reasoning, and this aid may not be the result of mere fortuitous coincidence whereby the content acts heuristically by cueing an accurate response. Rather, I would argue that the content of an argument, the context surrounding its consideration, and the activation of a reasoner’s belief systems may promote a different (or even an enhanced) type of reasoning than what is standardly observed when an abstract argument is resolved. In terms of methods, those employed by Wyer (1976), Markovits, Saelen, and Forgues (2009), and Stollstoff, Vartanian, and Goel (2012) have inspired a novel testing ground for the application of this integrative approach. This second stream of research offers methodological inspiration for my dissertation in that the role of an argument’s premises is highlighted.

**Content, Context, and Beliefs can Improve Analytical Strategies**

Most studies have shown that syllogistic content that activates one’s beliefs hurts reasoning or leads to accidental, heuristically-based improvements. Based on observations of card-sorting behavior, Wason & Shapiro (1971) drew opposite conclusions by suggesting that thematic content can improve reasoning in a non-heuristic manner by reducing the working memory load associated with analytic processing or encouraging a balanced weighting of the problem components. In particular, Wason and Shapiro compared two methods for presenting the parts of a reasoning problem – thematic problems referred to everyday experiences (e.g.,
Every time I go to Manchester, I go by train) whereas *abstract* problems did not (e.g., Every time there is an even number on the card, the other side has a vowel). The thematic condition was associated with greater accuracy rates and Wason and Shapiro speculated about the reasons behind this difference. Firstly, it could have been that concrete words (e.g., objects) are easier to remember and induce lower working memory loads than intangible ones (e.g., numbers and letters). Alternately, the fact that the relationship between words was concrete in the thematic condition may have made it easier to manipulate the problem components relative to the abstract condition in which the relationship between terms was again less concrete. Finally, it could have been that upon encountering 4 problem statements and being asked to derive a rule to describe them, that the 4 problem statements in the thematic condition represented relationships that were treated as equals (i.e., each statement is a trip to a given place using a given vehicle), whereas in the abstract condition it may have been easier to focus one’s attention on certain problem statements over others, thereby rendering an unbalanced judgment.

Wason and Shapiro proposed that everyday knowledge can improve the accuracy of reasoning by either facilitating working memory or ensuring that problem components are employed in a balanced fashion. Loken & Wyer (1983) also weighed in on the relationship between knowledge and reasoning by suggesting that everyday beliefs are built syllogistically. This would imply that everyday knowledge should not impede reasoning because syllogistic form is the natural context in which this content is routinely manipulated. Furthermore, Loken and Wyer showed that remembering one part of a syllogistically constructed belief system brings to mind the whole argument which again suggests that there is an intimate relationship between problem content and structure. In this case recalling content drives the recollection of an argument’s structure into which is embedded the remaining content. Loken and Wyer
hypothesized that when people encounter two premises in life (e.g., “smoking cigarettes at an early age leads to depending on cigarettes,” and “dependency on cigarettes leads to bad health”) they are implicitly pushed to form their own conclusion (A is C: smoking cigarettes at an early age leads to bad health). This newly learned belief is thought to be an ‘implicational molecule’ (Abelson & Reich, 1969, as cited in Loken & Wyer, 1983) such that when the belief is activated, the premises that originally generated it are also recollected. The inference process is thereby stored in memory as a single unit and recollection of one part of that argument brings to mind the remainder of it. This hypothesis received support through a study in which participants viewed and evaluated the plausibility of conclusion statements (e.g., A-C) that were preceded by syllogistically related (e.g., A-B, B-C) and unrelated (e.g., A-E, D-E) premises.

During a later memory test participants were equally successful at identifying familiar conclusions regardless of whether they had been preceded by syllogistically related or irrelevant. Yet, for false alarms, novel conclusions were substantially more likely to be accepted when they contained terms (A,C) that were previously linked together syllogistically. So, the novel conclusion C-A is apt to be mistaken for familiar because the A-B, B-C premises implied it whereas conclusion C-E is less likely to be falsely recognized because it was not primed by an earlier combination of syllogistically linked premises.

Piper (1985) talked about the role of context for reasoning from the developmental angle. Despite Piaget’s contention that abstract reasoning abilities come with age, Piper showed that when content is properly contextualized by reference to a particular type of narrative world (sci-Fi-abstract), pre-adolescents can reason about syllogisms adeptly and are in fact more successful than Grade 12 students who handle content best when it is contextualized within an everyday pragmatic or everyday contractual narrative. Piper had previously documented that Grade 6
students enjoy Sci-Fi narratives substantially more than Grade 12 students which suggests that when an argument is contextualized in a way that mimics a one’s everyday mode of thinking, that this can have an ameliorative effect on reasoning. Piper’s findings certainly challenge the Piagetian model of development, however this model may still hold true for reasoning outside of narrative contexts. What is key to reiterate however is that Piper’s study shows that reasoning is not isolated from context but greatly impacted by it. This study anticipates Falmagne and Gonsalves’ integrative approach and also illustrates their assertion that “propositions are defined in terms of possible worlds in which they are true” (p. 9). If the possible world to which the reasoner refers is a sci-Fi-abstract one in which anything goes, the propositions can be manipulated more easily in mind. As Wason and Johnson-Laird (1972) stated in their book *Psychology of Reasoning*, presenting a syllogism in the context of narrative makes the problem easier to “get into” because the reasoner is able to “project themselves by an act of imagination” (as cited in Piper, 1988, p. 191). Thus, the argument is not objective, it is “not a creature of constant hue, but chameleonlike,” in that it takes on the “colour of its surroundings” (p. 92).

**Premise Plausibility**

Methodological inspiration can firstly be credited to Wyer and Goldberg (1970) who, in an effort to demonstrate that syllogistic reasoning is actually probabilistic reasoning, compared participants’ reasoning about arguments with trusted and mistrusted major premises. Although they found no differences and the manipulation itself was not about the veracity of the content but rather the trustworthiness and prestige of the source, this study is nonetheless a touchstone for my research because it introduced the idea that participants’ perceptions of the premises may impact their evaluation of the conclusion.
Continuing along a similar vein, Wyer (1976) conducted a study that unsystematically considered what would happen to reasoning when one works with a major premise that is negative and implausible. Following from Wyer and Goldberg’s (1970) equation, $P_{AB} = P_A \times P_B$, Wyer assumed that under normal conditions the perceived plausibility of a conclusion would be derived from an equal weighting of the major and minor premise’s plausibility, $P_{\text{conclusion}} = P_{\text{premise1}} \times P_{\text{premise2}}$. Thus, a conclusion would be deemed 40% likely in the case of a major premise that was 50% likely and a minor premise that was 80% likely. Wyer wanted to determine whether manipulating the content in the major premise could distort this weighting process in a way that would either under-estimate or over-estimate the conclusion’s probability. For instance, one’s background experience could lead to the belief that “successful female movie stars are physically attractive” and the experimental task could make salient one of two beliefs, that “successful female movie stars are usually physically attractive” or “successful female movie stars are rarely physically attractive” (respectively consistent or inconsistent with previously held beliefs).

Participants were instructed to disregard previously held beliefs but the results of Wyer’s study showed that this was not possible, particularly in cases of negation involving the major premise. So, when the major premise made a negative affirmation (X was said to rarely or never belong to category Y) that was highly inconsistent with previously held beliefs, the plausibility of the conclusion tended to be underestimated. That is to say that when the major premise should have been trusted, it was not, and this meant that it contributed less to the equation than it should have. Much like Wyer’s materials (1976), those of Markovits and Nantel (1989) also contained premises that were at times implausible (e.g., “all things that are smoked are good for the health”) however this variation was once again unsystematically considered.
A systematic manipulation of the plausibility of major premises was first conducted by Markovits, Saelen and Forgues (2009) who employed a conditional reasoning paradigm to explore belief inhibition. They demonstrated that reasoners more successfully resolve belief-logic conflict in conclusions when arguments contain implausible (UBU) rather than plausible (BBU) premises. Interestingly, this effect was only evident for a certain class of arguments – specifically arguments in which premises took the Modus Ponens (MP) argument form (If P then Q, P is true) but not when they Affirmed the Consequent (AC) (if P then Q, Q is true). Markovits et al. interpreted the main effect in terms of inhibitory differences; in the implausible condition participants viewed an initially implausible statement (if something is a fruit, then it uses electricity) followed by a plausible minor premise (a tomato is a fruit) and an implausible conclusion (a tomato uses electricity). Given that participants had to accept the premises as true, the major premise necessitated inhibition of the assertion that “fruits do not use electricity” which seemingly eased the reasoner into accepting the conclusion that a “tomato uses electricity”.

In the plausible condition however, the major premise was believable (if something is a fruit, then it grows on trees) so inhibition was not immediately necessitated by the major premise. However, the minor premise (a tomato is a fruit) represented an atypical case for which P (fruit) is not Q (growing on trees), and directly contradicted the conclusion that a “tomato grows on trees”. The authors suggested that content inhibition was not necessitated by the premises because individually they were plausible. Thus, participants were hindered in their acceptance of the conclusion that a “tomato grows on trees,” because they had not engaged in pre-emptive inhibition of knowledge that would ultimately challenge this conclusion. Consequently, the plausible condition was more difficult than the implausible one.
A second study contrasted the same MP conditions and considered reasoners’ response times and general success (high, low) at solving conditional reasoning problems despite content-logic conflicts in the conclusion. They found that highly skilled reasoners took less time to evaluate conclusions preceded by implausible premises and were also more successful at resolving belief-logic conflict than if the conclusion had been preceded by plausible premises. On the other hand, participants with less skill at resolving belief-logic conflict took longer to evaluate conclusions preceded by implausible premises and were also less successful than if the belief-logic conflict in the conclusion had been premised by plausible propositions. The authors interpreted this interaction by suggesting that those who are able to quickly and accurately overcome belief-logic conflict in the conclusion are the ones who have “proactively inhibited contradictory information” (p. 119). This preventative strategy is wielded by reasoners who are capable of recognizing and decisively nipping in the bud, clear doubts about the major premise’s empirical value. Strong reasoners were less effective at resolving belief-logic conflict in the conclusion of arguments premised by believable content because they were not given grounds upon which to pre-emptively inhibit conclusion-contradicting content.

It is important to note that although Markovits and colleagues attempted to systematically compare reasoning based on plausible and implausible premises in their design; only the plausible condition contained what I will refer to as inter-premise content conflict. In other words, although plausible arguments contained two premises that at face value appeared to be empirically valid, in combination they contradicted each other. This inter-premise content conflict was not found in the implausible condition and therefore it could account for the observed differences in reasoning. To truly test the impact of premise believability on conclusion
evaluation future studies need to use believable premises that reinforce each other in the plausible as well as the implausible conditions.

Santamaria, Garcia-Madruga & Carretero (1996) wrote an article about the holistic conceptualization of arguments that applies to this criticism of Markovits and colleagues’ design. A key weakness that they pointed out about traditional syllogistic reasoning studies relates to the routine treatment of premises and content. As they explained, a syllogism is truly believable in cases where there is functional equivalence between what is meant pragmatically and what is asserted logically. In other words, the quantity and quality of the propositions must match participants’ assumptions about the relationship between concepts in the argument. For instance, to say that “some dogs are not cats” may approximate some everyday truth, yet because the statement does not universally disqualify all dogs from being cats, functional inequivalence renders the argument implausible. This insight guided my use of universal propositions when constructing plausible and implausible content.

Santamaria and colleagues also stressed the importance of thinking not only about whether a given proposition accords with everyday knowledge, but also whether, in combination, two premises seem reasonable or at odds with each other. The authors pointed out that there are instances where individually, premises seem plausible, yet holistically, they fail to project a reasonable picture of the everyday world. For example, an abstract problem with premises “Some C are A, and No B are A” is logically equivalent regardless of whether we embed content of the form, “Some students take examinations in September, and No bookworm takes examinations in September” or “Some students take examinations in September, and No teacher takes examinations in September”. Yet, in combination the premises result in arguments that are perceived as more or less realistic renderings of the everyday world. As a general rule, teachers
would be individuals who are perceived as less likely to be students than bookworms, therefore the first set of premises seems more at odds with everyday knowledge than does the second.

Following from this example, Santamaria and colleagues manipulated the relationship between concepts in otherwise logically equivalent syllogistic arguments, for instance by comparing conditions in which “students” were equated either with “teachers,” “blonds,” or “bookworms”. Each case was logically comparable yet participants’ reasoning seemed to be guided by their implicit assumptions, for instance that “teachers” usually aren’t “students,” whereas “bookworms” are nearly always studying. In a nutshell, Santamaria and colleagues showed that reasoning is driven by the plausibility of the premises as a whole rather than by the individual plausibility of the components. As such, two problems with the same logical structure and seemingly plausible premises may be solved with differing degrees of success due to the fact that their premises are holistically more or less consistent with everyday knowledge.

The relationship between premises, as highlighted by Santamaria and colleagues is relevant to Markovits and colleagues’ plausible MP condition because in this case there was a conflict between the major and minor premises that prevented pre-emptive content inhibition. One can speculate that premises must present an uncompromising conflict if inhibition is to be inspired; if however the premises are independently plausible yet holistically implausible ambiguity results and inhibition is not inspired. In fact, even if a reasoner wanted to inhibit something it would be quite effortful for them to come up with an appropriate proposition to reject since the statements, individually considered, are independently plausible.

The application of an inhibition mechanism to reasoning was recently elaborated by Stollstoff, Vartanian and Goel (2012) who sought to extend what is known about right lateral prefrontal cortex (rLPFC) activity and conflict mediation – is rLPFC activity associated
specifically with belief-logic conflicts or more generally with belief-content conflicts? Stollstorff et al. administered three-term spatial relational arguments to participants. Behaviorally speaking participants took longer but were no less accurate when evaluating syllogisms containing belief-content conflicts within their premises compared to neutral premises. Notably, what was also observed that the rLPFC was active when participants processed premises that did not accord with everyday knowledge (belief-content conflict) albeit that incongruence between beliefs and structure was linked with a higher level of rLPFC activation. Thus, given the rLPFC’s role in conflict mediation it can be concluded that the mere presence of implausible or “counterfactual” content is sufficient to inspire inhibition, yet the rLPFC is most active in cases where one’s everyday beliefs are in conflict with the content of premises in addition to the structure of the argument. In either case, the authors point out that inhibition is necessary because both belief-content and belief-logic conflict necessitate “decontextualization – a separation between previous knowledge and the information held in working memory” (p. 28) for task completion.

**Purpose of Current Studies**

So far we have established that an implausible major premise may help reasoners avoid belief bias through the inhibition of competing beliefs, a function that may stem from rLPFC activity. Given that these relationships have only been tested using conditional reasoning and spatial relational arguments, the over-arching goal of my research is to explore the role of premise believability in syllogistic reasoning. I would anticipate that another contribution that may come out of my research is a critical commentary about the 6 reasoning models that were previously described. These models are in fact rather limited as soon as one starts to think about the empirical validity of premises and their ensuing effects on reasoning.
Based on a review of the models it seems to me that the least useful are those which consider reasoning to be a backwards process of conclusion evaluation (e.g., selective scrutiny). The most useful would be the mental models and verbal reasoning accounts which view reasoning as happening in a forwards manner. Although Oakhill, Johnson-Laird and Garnham (1989) and Polk and Newell (1995) conceive of reasoning in an order where premises are considered before conclusions, they did not explicitly consider what happens to reasoning when these premises are plausible and implausible. Thus, although I will reiterate that my goal is not to evaluate the models, I would expect that the series of studies I will report will help to develop an integrated perspective about reasoning and the belief bias effect, as well as suggesting future directions of study highlighting the role of belief-content conflict and premise-based reasoning within these models.

Please refer to Table 4 which outlines the key goals of Studies 1A, 1B, 2, and 3. In Study 1A I endeavor to determine whether belief-content conflict in the major or minor premise attenuates or aggravates belief bias stemming from belief-logic conflict in the conclusion. In particular, I explore how belief-content conflict within valid and invalid syllogisms may impact reasoning accuracy, response time and confidence. In Study 1B I aim to establish whether a proposition is perceived as being equally likely or unlikely when evaluated in isolation or in the mixed company of two other propositions. One of the key questions addressed in Study 1B is whether or not belief-content conflict in the premises transforms participants’ perceptions of the conclusion? This study really captures the spirit of the integrative approach by looking at the role of contextualization and treating the propositions of an argument as interactive and potentially mutually constitutive. Study 1B also aims to compare participants’ interest in content and their orientation towards structural elements of syllogistic arguments. If belief-content conflict
attenuates belief bias does it do so by inhibiting participants’ interest in content and/or encouraging a focus on structural elements of the argument?

Table 4

Predictions per Study

<table>
<thead>
<tr>
<th>Study</th>
<th>Key predictions</th>
<th>Supported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Premise-based belief-content conflict (BCC) attenuates belief bias in <strong>incongruent valid</strong> and <strong>congruent invalid</strong> trials.</td>
<td>Yes</td>
</tr>
<tr>
<td>1B</td>
<td>BCC diminishes participants’ sensitivity to conclusion content. BCC encourages evaluation of syllogistic structure.</td>
<td>Yes No</td>
</tr>
<tr>
<td>2</td>
<td>Major premise-based BCC attenuates belief bias in <strong>incongruent valid</strong> and <strong>incongruent invalid</strong> trials. <strong>Difficulty</strong> of syllogistic structure limits the efficacy of BCC as an attenuator of belief bias.</td>
<td>Yes Partially</td>
</tr>
<tr>
<td>3</td>
<td>Major premise-based BCC attenuates belief bias in <strong>incongruent valid</strong> and <strong>incongruent invalid</strong> trials. <strong>Difficulty</strong> of syllogistic structure limits the efficacy of BCC as an attenuator of belief bias. BCC reduces participants’ time spent evaluating conclusion content.</td>
<td>Yes Partially Yes</td>
</tr>
</tbody>
</table>

There were three main goals of Study 2. Firstly, to replicate the findings reported in Study 1. Secondly, to introduce syllogistic difficulty level as a new independent variable and to observe how it may shape the outcome of syllogistic reasoning following belief-content conflict. Lastly, Study 2 also investigates a new set of semantic conditions for invalid syllogisms and tests whether belief-content conflict can promote accurate reasoning about invalid syllogisms that contain plausible conclusions. Study 3 attempts to confirm the findings reported in Study 2. It also investigates the amount of time that participants devote to the evaluation of plausibility of
syllogistic arguments that vary in terms of belief-content conflict and difficulty level. A key question that is being addressed in Study 3 is whether the difficulty level of the syllogistic argument and/or the pattern by which semantic content is embedded in the argument impacts participants’ likelihood of focusing on the empirical validity of the conclusion.
Beliefs have been shown to bias syllogistic reasoning (e.g., Evans, Barston, & Pollard, 1983). In the case of congruent trials in which the structure and content of a syllogism match, (e.g., valid syllogism with a believable conclusion), accuracy rates tend to be substantially higher than in incongruent trials where the structure and content are at odds (e.g., valid syllogism and unbelievable conclusion). Introducing belief-content conflict in the major premise of conditional reasoning problems has been shown to alleviate belief bias in strong reasoners (Markovits, Saelen, & Forgues, 2009), thus the purpose of Study 1A is to determine whether belief-content conflict can similarly attenuate belief bias in syllogistic reasoning tasks. Building on Markovits and colleagues’ work, this study investigates not only the potential impact of belief-content conflict in the major premise, but also how belief-content conflict in the minor premise may contribute to conclusion evaluation.

I expect that premise believability [(B)elievable or (U)nbelievable] will impact reasoning about incongruent valid trials. When participants encounter an unbelievable premise they should be less vulnerable to belief-logic conflict in the conclusion. Thus, belief-content conflict at the beginning of an argument should attenuate belief bias when reasoners judge the structure of incongruent, valid syllogisms. This hypothesis accords with Markovits et al.’s finding that an implausible major premise promotes accurate resolution of belief-logic conflict for modus ponens inferences in a conditional reasoning paradigm, presumably because such a belief-content conflict cue prompts pre-emptive content blocking.
I will investigate reasoning about valid syllogisms with (a) three believable propositions (BBB), (b) believable premises and an unbelievable conclusion (BBU), and (c) one unbelievable premise and an unbelievable conclusion (UBU or BUU). The first condition (BBB), being entirely congruent, should promote high rates of accuracy for participants’ validity judgments. The incongruent second condition (BBU) should prompt the standard belief bias effect which should substantially diminish accuracy. The third condition (UBU or BUU), consisting of belief-content conflict that may serve as a sort of ‘warning signal,’ should attenuate belief bias and lead to accuracy that is significantly higher than that observed in the BBU condition.

I will also compare participants’ accuracy when evaluating invalid syllogisms of the forms BBB, BBU, and UBU/BUU. Invalid syllogisms with 3 believable propositions (BBB) are incongruent and participants should evaluate them with poor success. Invalid syllogisms with believable premises and an unbelievable conclusion (BBU) are congruent and should be associated with inflated accuracy levels. The condition that I have constructed in which an unbelievable conclusion is preceded by belief-content conflict (UBU/BUU) is congruent in that invalid structure matches the conclusion’s unbelievable content, yet the presence of belief-content conflict in the premises may translate into lower accuracy than is observed in the BBU condition. There are a variety of reasons why this might happen; belief-content conflict may cause System 1 to be inhibited thereby disrupting content-oriented processing that would lead to heuristically inflated accuracy; it may heighten epistemic vigilance and encourage System 2; or it could even prompt System 2 to shift from one type of analytic strategy to another.

The amount of time expended during reasoning and participants’ confidence about their reasoning will be compared between conditions. These comparisons will help to determine the type of processing that is promoted by belief-content conflict. If for instance belief-content
conflict promotes System 2 processing then elaborated response times would be expected in the third condition (UBU/BUU). It is not possible to make a clear-cut hypothesis about variations in confidence ratings given that there is debate in the literature about whether confidence reflects participants’ analytical performance or is generated heuristically (e.g., Shynkaruk & Thompson, 2006).

Methods

Participants

Sixty-two University of Toronto, Scarborough Campus (UTSC) undergraduate students (26 males, 36 females) participated in this study. They received credit towards their final course grade in Introductory Psychology in exchange for their involvement with this study. Recruitment was limited to native English speakers and to those who began speaking English by the age of 3 years. Students were also asked to indicate if they had received formal training in logic (e.g., undergraduate courses such as PSYD56H3: Creativity, Reasoning and Problem Solving; PHYLB50H3: Symbolic Logic I). None of the participants reported a background in reasoning or symbolic logic.¹

Materials

Pre-testing with UTSC undergraduates prior to Study 1 helped to identify the average difficulty level of numerous syllogisms. The results of this pre-testing guided the syllogism selection process for the current study. Six syllogisms varying in structure (valid; invalid), content (non-semantic; semantic) and difficulty level (easy; difficult) were presented as warm-up trials (see Appendix A). Eight syllogisms varying in structure (valid; invalid) and major premise

¹ Given that these courses often equip students with shortcuts so that they do not need to engage in reasoning in order to determine whether an argument is valid or invalid, it was important to avoid recruiting individuals with this background so that our observation of naturally occurring reasoning and the belief bias effect could be maximized.
proposition type (universal, for example “all A are B”; particular, for example “some A are B”) were included as non-semantic test trials. Each combination (e.g., universal-valid; particular-valid; universal-invalid; particular-invalid) was represented by two different syllogisms which were administered within-subjects (see Appendix B). Twenty-four syllogisms varying in structure (valid; invalid), major premise proposition type (universal, all; particular, some), and believability (all propositions believable, BBB; conclusion unbelievable, BBU; universal proposition unbelievable, UBU or BUU) were included as semantic test trials. Based on the design there were 12 possible combinations. Each combination was represented by two different syllogisms which were administered within-subjects (see Appendix C for the design of Study 1A and Appendix D for the content of the syllogisms).

Measures

During the test trials the dependent measures were: accuracy (0 = miss; 1 = hit) of participants’ validity judgments, the amount of time (msec) spent reasoning about the conclusion and judging the argument’s validity, and participants’ reported confidence about the validity judgment on a scale of 1 to 5 (1 = not at all confident; 5 = extremely confident).

Procedure

Participants were seated at individual computer terminals and were instructed to turn off their cellular phones. They were encouraged to take their time reading the instructions and to press the space bar to advance at their own pace through the written instructions. The trials were programmed using E-Prime. Participants were introduced to the syllogistic task by way of written instructions that appeared on the computer screen. Firstly, participants were introduced to the structural components of the logical syllogism by being told that “the logical syllogism is a form of deductive argument with three parts: the minor premise, the major premise, and the
conclusion.” They were provided with one example of an abstract (non-semantic) syllogism and a second example depicting a concrete (semantic) syllogism. Participants were then introduced to the validity judgment task; they were told that their “job would be to assume that the major and minor premises are true,” and they were asked “if this is the case, then is the conclusion true or false? Is the conclusion necessary given that the premises are true?” Participants were informed about how to use the keyboard to record their responses. They were asked to press “z” if they thought the conclusion was valid and “followed necessarily from the premises” or “m” if they thought that the argument was invalid and “did not follow necessarily from the premises”.

Participants completed warm-up trials before moving along to the test trials. During each trial participants had unlimited time to view the major and minor premises, advancing to the next screen at will by pressing the space bar. When the conclusion statement was presented, the premises also reappeared so that participants did not have to remember the parts of the argument. Participants had unlimited time to judge whether or not the conclusion statement followed necessarily from the premises, and to indicate their confidence in this judgment, a response that was prompted on a separate screen and required them to press a number from 1 to 5 on the keyboard. Appendix E provides a mock-up of the display screens used in this experiment.

During the warm-up trials participants judged whether each syllogism was true (valid) or false (invalid). An accuracy score per warm-up trial was generated. Participants with below-average aptitude for syllogistic reasoning were identified by referring to individual accuracy for the easy warm-up trials. The easy warm-up trials consisted of syllogisms that were solved accurately by more than 90% of participants during pre-testing. Participants in the current study who failed to solve the easy trials were excluded from further analyses. Additionally, participants who solved the easy warm-up trials correctly but scored at chance levels for the remaining
practice trials were only kept in the sample if they scored above chance on the non-semantic test trials. Based on these criteria 5 participants out of 62 were excluded from data analysis.

**Design**

Gender was a between-subjects variable. The within-subjects variables, administered with repeated measures, included structure (valid; invalid), major premise proposition type (universal, *all*; particular, *some*), and believability (all propositions believable, BBB; conclusion unbelievable, BBU; universal proposition unbelievable, UBU or BUU). The non-semantic and semantic test syllogisms were organized into separate blocks of trials. For Order 1 the blocks were each randomly ordered and the non-semantic block was presented prior to the semantic block. For Order 2 the order within each block was reversed and the non-semantic block was again presented prior to the semantic block. Orders 3 and 4 saw the semantic block presented prior to the non-semantic block; the orders within the blocks were identical to those in Order 1 and Order 2, respectively.

**Results**

**Manipulation Checks**

During pre-testing 28 participants used a 7-point rating scale to evaluate the believability of the 72 propositions making up the test syllogisms that were used in the current study. The average plausibility of a believable (B) proposition was 89%; while the average plausibility of an unbelievable (U) proposition was only 29%. This difference was significant ($p < .0001$) and consistent across the conditions in the study. In other words, the syllogisms used as stimuli in this study were comprised of comparably plausible B propositions and implausible U propositions. The average plausibility of each syllogism in the study was calculated by adding up the participants’ ratings per proposition. Each proposition had the potential of receiving 7 points
(highly believable) and each proposition was rated 28 times. The overall plausibility of a given syllogism is therefore calculated by summing up the 28 ratings of the syllogism’s 3 propositions and dividing that number by the total possible points (588) that the argument could have obtained if it were 100% plausible. The equation is as follows: (Sum of ratings from Major + Minor + Conclusion) / 588 x 100%.

As anticipated, the plausibility of the arguments differed systematically according to the categories (BBB, BBU, UBU/BUU) of the Believability condition. Arguments with three believable propositions (BBB) had an average plausibility rating of 88%, compared to a 68% plausibility rating of arguments with believable premises and unbelievable conclusions (BBU), and a 50% plausibility rating for arguments with unbelievable major premises and conclusions (UBU or BUU). Believability is a key manipulation in Study 1A so it is important to reiterate that (a) believable premises were 60% more plausible than unbelievable premises, (b) the stimulus set consisted of believable propositions that were equally plausible and unbelievable propositions that were equally implausible, and (c) the believability conditions were marked by semantic differences in perceived plausibility.

*The Valid Syllogisms*

A repeated measures analysis of variance tested for main effects and interactions between participants’ gender, and the two key structural variables associated with the valid syllogisms: believability of the content (BBB, BBU, UBU/BUU) and the major premise’s form (universal, “all”; particular, “some”) for participants’ accuracy, the time spent evaluating the syllogisms, and participants’ self-reported confidence about their validity judgments. Please refer to Tables 5 and 6 for the main effects of believability and major premise on these variables for valid trials. Appendix I summarizes non-significant results related to Believability.
Table 5

*Main Effects of Believability on Accuracy, RT, and Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Believability</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU/BUU</th>
<th>p &lt;</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>.91</td>
<td>.43</td>
<td>.67</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Reaction Time (ms)</strong></td>
<td>14674.43</td>
<td>22101.47</td>
<td>19759.83</td>
<td>.001</td>
<td>*</td>
</tr>
<tr>
<td><strong>Confidence</strong></td>
<td>4.10</td>
<td>3.48</td>
<td>3.64</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Mauchly’s test of sphericity showed that the variances were significantly different (p<.05). The Greenhouse-Geisser degrees of freedom and p-value were referred to.*

**Accuracy.** Participants’ gender and the form of the major premise did not independently impact task accuracy in a way that would distinguish between male and female participants, or universal and particular statements. The interaction between these two variables was insignificant, as was the two-way interaction between believability and Gender, and the three-way interaction between major premise, believability and Gender.

Table 6

*Main Effects of Major Premise on RT and Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Major Premise</th>
<th>Universal</th>
<th>Particular</th>
<th>p &lt;</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reaction Time (ms)</strong></td>
<td>17388.42</td>
<td>20302.07</td>
<td>.01</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Confidence</strong></td>
<td>3.83</td>
<td>3.65</td>
<td>.01</td>
<td>ns</td>
</tr>
</tbody>
</table>
Believability had a significant effect on accuracy, $F(2, 110) = 60.44$, $p<.001$, $\eta^2 = .19$ (see Figure 1). Post hoc testing by way of paired $t$ tests with a Bonferonni correction showed that all three levels of this variable were unique. Specifically, accuracy was highest in the BBB condition where valid structure was congruent with believable content and lowest in the BBU condition where valid structure conflicted with unbelievable content. Accuracy in the UBU/BUU condition was found to be significantly higher than in the incongruent BBU condition, although it was not on par with that observed in the BBB condition.

Figure 1

*Main Effect of Believability on Accuracy for Valid Syllogisms*
There was a significant interaction between major premise and believability for accuracy, $F(2, 110) = 7.41$, $p < .01$ (see Figure 2). The effect size was small, $\eta^2 = .02$. Post hoc tests revealed that the effect of the UBU/BUU condition on accuracy was higher when the major premise was a particular statement than when it was a universal statement. Given that the universal proposition was used to present unbelievable content in this subset of trials, this suggests that the unbelievable content is somewhat more successful at boosting accuracy when it is deployed as the minor premise (BUU) than when it appears as the major premise (UBU).

Figure 2

*Interaction Effect of Believability x Major Premise on Accuracy for Valid Syllogisms*

*Reaction time.* The amount of time participants spent reasoning about whether an argument’s conclusion followed necessarily from its premises was no different when the interactions between believability and major premise, major premise and Gender, and
believability and Gender were considered. Additionally, male and female participants spent the same amount of time evaluating the conclusion.

There were significant main effects for believability, $F(2,110) = 16.16, p<.001, \eta^2 = .05$, (see Figure 3), and major premise type, $F(1, 110) = 10.33, p<.01, \eta^2 = .01$ (see Figure 4).

Figure 3

*Main Effect of Believability on Reaction Time for Valid Syllogisms*

![Reaction Time Graph](image)

Post hoc testing comprised of paired $t$ testing and the application of Bonferroni corrections revealed that participants spent significantly more time reasoning about incongruent syllogisms of the form BBU and UBU/BUU compared to congruent syllogisms (BBB).
Participants also spent significantly longer reasoning about syllogisms that opened by making a particular assertion (e.g., some A are B) as compared to a universal assertion (e.g., all A are B). If we recall that a universal proposition was used to present unbelievable content in this experiment, then this suggests that participants would by default be likely to spend more time considering belief-content conflict when such a cue appeared as the minor rather than the major premise.

Confidence. The interaction between major premise type and believability was not significant, nor was there a significant effect of Gender, either alone or in combination with the other independent variables. Participants were significantly more confident that they had
accurately evaluated the structure of arguments that opened with a universal as opposed to a particular assertion, $F(1, 110) = 7.89, p < .01, \eta^2 = .01$ (see Figure 5).

Figure 5

*Main Effect of Major Premise on Confidence for Valid Syllogisms*

Confidence level was also significantly impacted by the believability of the argument, $F(2, 110) = 11.91, p < .001, \eta^2 = .08$ (see Figure 6). Post hoc testing revealed that confidence was highest in the congruent condition, BBB, in which the conclusion was plausible and the structure was valid. Participants were significantly less confident about the accuracy of their judgments for syllogisms of the form BBU and UBU/BUU. The difference between these latter two conditions is negligible.
The Invalid Syllogisms

A repeated measures analysis of variance was employed to determine if there were any main effects or interactions between participants’ gender, and the two key structural variables associated with the invalid syllogisms: believability of the content (BBB, BBU, UBU/BUU) and the major premise’s form (universal, “all”; particular, “some”) for participants’ accuracy, the length of time participants spent evaluating the syllogisms, and participants’ self-reported confidence about the validity judgments they made. Please refer to Table 7 for the main effects of Believability on these dependent variables for invalid syllogisms.
Table 7

*Main Effects of Believability on Accuracy, RT, and Confidence for Invalid Syllogisms (Study 1A)*

<table>
<thead>
<tr>
<th>Believability</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU/BUU</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.30</td>
<td>.68</td>
<td>.62</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>19285.32</td>
<td>22238.28</td>
<td>18244.85</td>
<td>.05</td>
<td>*</td>
</tr>
<tr>
<td>Confidence</td>
<td>3.86</td>
<td>3.77</td>
<td>3.77</td>
<td>.05</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Mauchly’s test of sphericity showed that the variances were significantly different ($p<.05$). The Greenhouse-Geisser degrees of freedom and $p$-value were referred to.

**Accuracy.** There was no main effect of Major Premise type or Gender. Believability had a significant main effect on accuracy, $F(2, 110) = 61.22, p < .001, \eta^2 = .15$ (see Figure 7).

Figure 7

*Main Effect of Believability on Accuracy for Invalid Syllogisms*
Believability also interacted significantly with Major Premise type, $F(2, 110) = 9.96, p < .001, \eta^2 = .02$ (see Figure 8).

Figure 8

*Interaction Effect of Believability x Major Premise on Accuracy for Invalid Syllogisms*

Post hoc testing showed that overall, participants were the least accurate when evaluating invalid syllogisms that were of the incongruent variety, BBB. This type of syllogism was accurately judged to be invalid only 30% of the time. On the other hand, participants were significantly more accurate when evaluating invalid syllogisms of the form BBU (68%), and UBU/BUU (62%). The difference between the BBU and UBU/BUU conditions was not significant on its own, however taking Major Premise type into consideration revealed an interesting interaction.

Post hoc testing revealed the following about the interaction between Believability and Major Premise type for participants’ accuracy when evaluating invalid syllogisms: arguments with a universal major premise (All A are B) were most accurately judged in the congruent
condition, BBU, and the least accurately judged in the incongruent condition, BBB. Interestingly, the ‘warning signal’ condition, UBU, was associated with an accuracy level that was significantly lower than that observed in the BBU condition, yet significantly better than the accuracy level observed when there was a high degree of incongruency in the syllogism, BBB.

Invalid syllogisms that began with a particular statement, (meaning that the implausible universal statement was held until the second step of the argument in the BUU condition), showed the same difference in accuracy between the BBU and the BBB conditions, specifically the congruent BBU condition was associated with high accuracy and the incongruent BBB condition was associated with low accuracy. However, no difference between the congruent BBU and warning signal BUU conditions was observed. The accuracy levels associated with the two warning signal conditions, UBU and BUU did not differ significantly from each other. Regardless of whether the universal warning signal appeared as the major premise or minor premise of the argument, participants were moderately accurate (58-66%).

What is it that the belief-content conflict accomplishes as a major premise? It seems to change conclusion evaluation so that the accuracy outcomes for the UBU and BBU conditions are substantially different. When the major premise is universal and plausible and the conclusion is implausible, belief bias proliferates and accuracy is quite high. When the implausible conclusion is preceded by a major premise that is universal and implausible, this artificially inflated accuracy is reined in and accuracy diminishes to 66%. This level of accuracy is substantially greater than chance (p<.05) but not on par with the 80% range. Thus it seems unlikely that participants are guessing or responding heuristically according to syllogistic content which would artificially boost their accuracy.
What happens when belief-content conflict falls in the minor premise? It seems that there is no difference between the BUU and BBU conditions. In syllogisms with an implausible conclusion, *particular and plausible* major premise, and *universal and plausible minor premise*, accuracy is approximately 60% which suggests that belief bias is occurring since this is substantially better than chance performance ($p<.05$) and a marked improvement relative to the incongruent BBB condition where beliefs work against logical structure and cue inaccurate responses. When the implausible conclusion is preceded by a major premise that is *particular and plausible* and a minor premise that is *universal and implausible*, accuracy does not change relative to the BBU condition, remaining in the 60% range. If belief-content conflict in the minor premise were a successful warning signal, accuracy should be lower in the BUU condition than in the congruent BBU condition because the artificial inflation afforded to accuracy by semantic and structural symmetry should have been reduced. This lack of change therefore indicates that belief content conflict in the minor premise may have been an ineffective warning signal. Alternately, it could be that the particular major premise of the BBU condition failed to induce a strong belief bias effect to then be countered by the warning signal afforded by belief-content conflict in the minor premise. The content of arguments starting with a strong universal statement may be perceived as more trustworthy than those starting with a less dominant particular statement. As such, beliefs may be less biasing for arguments of the form IAI compared to arguments of the form AII. Thus, perhaps it is not that belief-content conflict in the minor premise less effectively moderates belief-logic conflict in the conclusion, but rather, that there is less of a belief bias effect to be moderated when an argument opens with a particular statement.
Reaction time. There was no significant main effect of Gender or Major Premise, nor did these variables interact in a way that significantly impacted reaction time. The interactions between Believability, Major Premise, and Gender, were also non-significant. There was a significant, albeit small ($\eta^2 = .01$) effect of Believability on reaction time, $F(2, 110) = 3.63$, $p<.05$ (see Figure 9). Post hoc paired samples t-testing revealed that participants took significantly longer to evaluate invalid syllogisms with plausible content throughout (BBB) or plausible premises in combination with an implausible conclusion (BBU) than syllogisms with implausible content in the preliminary and concluding parts of the argument (UBU/BUU).

Figure 9
Main Effect of Believability on Reaction Time for Invalid Syllogisms
Confidence. The interaction between Believability and Major Premise impacted participants’ self-reported confidence, $F(2, 110) = 5.45, p<.01, \eta^2 = .02$ (see Figure 10).

Figure 10

*Interaction Effect of Believability x Major Premise on Confidence for Invalid Syllogisms*

Post hoc testing revealed that incongruent syllogisms of the form BBB were judged with significantly more confidence when they started with a universal statement as opposed to a particular statement. Furthermore, confidence across the levels of the Believability variable differed significantly for arguments with major premises that were universal – the incongruent condition, BBB, was judged with great confidence whereas the congruent condition, BBU, and the ‘warning signal’ condition, UBU, were evaluated with significantly less confidence. Although the incongruent condition, BBB, stood out when the invalid syllogism had a universal proposition as its major premise, there was no such distinction for invalid syllogisms with a particular major premise (Some A are B). The levels of the Believability variable were
comparable to one another; invalid syllogisms with a major premise that was particular in nature were judged with a moderate level of confidence by participants regardless of the form of their semantic content (incongruent, congruent, ‘warning signal’). Between the Major Premise conditions, the congruent BBU and ‘warning signal’ UBU/BUU semantic forms did not differ from each other substantially.

*Does the ‘Warning Signal’ Effect Translate to Ensuing Trials?*

This post-hoc question was answered by comparing participants’ accuracy on incongruent trials (valid: BBU; invalid: BBB) when said trials were preceded by either a ‘neutral’ syllogism or a syllogism containing a ‘warning signal’. For the valid trial type, a syllogism of the form BBB was deemed to be neutral as compared to a syllogism of the form UBU, which contained an implausible major premise. Earlier analyses showed that the UBU form of valid syllogism was evaluated more accurately than the BBU form, suggesting that the implausible major premise acted as a semantic marker leading to more critical assessment of the ensuing and incongruent conclusion. For the invalid trial type, a non-semantic syllogism served as the ‘neutral’ counterpart for comparison with the ‘warning signal’ syllogism of the form BUU. Both preceded the BBB form of syllogism which was incongruent since the underlying structure of the argument was invalid.²

Accuracy was compared for incongruent trials that were and were not preceded by a syllogism containing a warning signal premise. Two stimulus orders were used to construct the conditions; this ensured that each combination (non-warning signal trial precedes incongruent

---

² The question of whether carry-over effects occur was raised by Dr. Moscovitch. This question was then answered in a post-hoc fashion relying on the available sequence of trials that had been previously administered. As such, it was not possible to make a perfectly clean comparison of valid and invalid trials because the potential impact of belief-content conflict was considered in the context of semantic neutral trials in the first instance and non-semantic neutral trials in the latter instance. Future studies should look for carry-over effects in instances where valid and invalid arguments containing belief-content conflict systematically precede both semantic neutral and non-semantic neutral trials.
trial; warning signal trial precedes incongruent trial) consisted of an early and a late trial. This prevented practice effects from confounding the comparison; a problem that could have cropped up if for instance the warning signal trials were also the trials that participants solved towards the end of the experimental session compared to innocuous trials that were perhaps encountered at the beginning of the session. No significant differences in accuracy were observed which suggests that a warning signal (implausible content) improves performance on an individual trial but does not impact the outcome of reasoning on ensuing trials. It seems as though the average reasoner treated syllogistic trials as independent entities during the experiment.

Summary of Findings and Preliminary Interpretations

Valid syllogisms. For valid syllogisms participants successfully judged the structure of congruent syllogisms 91% of the time. This condition replicates previous work which demonstrates what might be termed an ‘artificial amelioration’ of syllogistic reasoning ability, an effect that stems from the fact that the content of the argument just happens to match its structure. When the conclusion was implausible accuracy plummeted to 43% (significantly below the chance level of 50%, p<.05), suggesting that participants were not guessing but were instead making a deliberate, yet misguided judgment. This condition reiterates the classically documented effect of beliefs on syllogistic reasoning about incongruent arguments. The so-called ‘warning signal’ condition, consisting of belief-content conflict in the major or minor premise seemed to attenuate belief bias stemming from belief-logic conflict in the conclusion. Average accuracy on belief-content conflict trials was 67%. An interaction between Major Premise and Believability shed further light on the relationship between belief-content conflict and accuracy, showing that this warning signal was more successful in the minor premise (BUU) than in the
This finding might be explained as a recency effect\textsuperscript{3} whereby belief-content conflict is more effective in the minor than the major premise because it is located at the problem’s end and therefore more likely to be foremost in the participant’s mind when they encounter the conclusion.

The reaction time analyses showed that although participants may have been more accurate when evaluating incongruent conclusions that were preceded by a warning signal, they did not take longer to complete this evaluation when compared to the traditional incongruent trial type in which the implausible conclusion was preceded by believable premises. Not surprisingly, the mere presence of incongruent content entailed a longer period of consideration than did the presence of entirely plausible and congruent semantic material, as was the case in valid syllogisms of the form BBB. Finally, upon considering the confidence with which participants evaluated the syllogisms, confidence was highest for the congruent condition and significantly lower for syllogisms containing incongruent content, regardless of the presence or absence of belief-content conflict at the beginning of the argument.

These findings suggest several things about the way valid arguments are evaluated. Participants tackled arguments differently depending on whether they agreed or disagreed with the premises. In the case of the BBU condition, participants were in essence led down the garden path; they read plausible major and minor premises that were easy to endorse before being presented with an implausible conclusion. Participants spent about 22 seconds attempting to determine whether this unbelievable conclusion followed necessarily from the premises. In the end their accuracy was low, below mere chance levels, yet they felt relatively confident, approximately 70\% certain, that they had accurately evaluated the argument. This combination of low accuracy and moderately high confidence suggests that participants were prompted

\textsuperscript{3} As suggested by Dr. Vartanian.
heuristically by the conclusion’s implausible content to declare that the argument was invalid. In
the case of the ‘warning signal’ condition, participants were presented with an implausible
premise at the outset of the trial. Their accuracy was significantly improved yet their response
time and confidence were on par with that of the BBU condition. This suggests that the warning
signal did not push participants to think more, but rather to think differently.

Analytic strategies that might be at play. It is possible that belief bias was attenuated by
belief-content conflict in the premises (UBU/BUU) because this warning signal induced (a)
forwards (premise-based) as opposed to backwards (conclusion-based) reasoning, and/or (b)
hypothetical and complex (multi-dimensional) relational as opposed to pragmatic and basic
(binary) relational thinking. By virtue of the fact that an implausible proposition is at odds with
one’s everyday knowledge, belief-content conflict at the outset of an argument is likely to attract
reasoners’ attention so that they engage in a forwards direction of reasoning, one that emphasizes
the premises and casts them as the foundations for the problem. Besides encouraging forwards
reasoning, belief-content conflict in the premises may inspire participants to imagine
hypothetical worlds in which “unlikely” events could be true as opposed to denying them
because they are “impossible” in the real world.

It would be easier to build this type of imaginary world early into the reasoning process
than at the end because at the end of a problem one would have to disqualify the real world they
had referred to to that point in order to make room for a new and imaginary world that might
accommodate an implausible conclusion. The presence of belief-content conflict in the premises
of an argument might push participants to ‘willingly suspend their disbelief’ by imagining a
context that would permit the implausible, an idea that accords with Piper’s (1985)
demonstration that despite their age, pre-adolescents are successful syllogistic reasoners so long
as the problems are framed appropriately within a narrative science fiction world. This willingness to imaginatively engage with problem content could help to improve accuracy despite belief-logic conflict in the conclusion by freeing a reasoner from deliberating about impossible (!) content, removing the pressure of denying the implausible, and encouraging the application of logical strategies.¹

Invalid syllogisms. Participants were poor at solving incongruent syllogisms of the form BBB. Their average accuracy was 30%, much below chance levels (p<.05), and their average confidence was moderately high – 77%. These two findings suggest that participants were not guessing and instead were making a deliberate judgment that was guided by the semantic content of the syllogism’s conclusion. In a nutshell, the believable conclusion afforded participants a false sense of security. When the believability of the conclusion matched the structure of the argument, (BBU), participants were less confident yet more accurate in their judgments of the arguments’ structure. In other words, participants often denied implausible content upon encountering it in the conclusion, but at the same time felt less confident that this strategy was an effective one. When belief-content conflict was presented before the implausible conclusion, participants were as successful and as confident about their judgment as when there was no warning signal and the conclusion was congruent. Yet, they took less time to make this judgment. Believability and Major Premise type interacted and significantly impacted accuracy and participants’ self-reported confidence. Accuracy for congruent trials (BBU) was higher for arguments that began with a universal major premise than a particular one. Belief-content conflict only diminished the belief bias effect associated with the congruent trial when it appeared as the major premise. For trials in which the belief-content conflict was embedded in
the minor premise, (i.e., the major premise was particular), accuracy did not diminish from the congruent to the warning signal trial.

The interaction between Believability and Major Premise type yielded different effects on accuracy for valid and invalid syllogisms with belief-content conflict serving as a better cue in the minor premise for valid trials and the major premise for invalid ones. At first glance the interactions seem contradictory; however closer inspection suggests otherwise. Valid trials present belief-content conflict followed by belief-logic conflict; invalid trials present belief-content conflict followed by belief-logic agreement. Morsanyi and Handley (2012) suggested that participants are intuitively aware of syllogistic structure which suggests that they are operating under different pressures in both trial types. The belief-content conflict cue in valid trials operates in a situation where there is great pressure for resolution given the presence of belief-logic conflict in the conclusion. It could be that having the belief-content conflict located in close proximity to the belief-logic conflict ups the ante and spurs on System 2 reasoning. After all, a high degree of conflict stemming from syllogistic structure, content, and beliefs, would be aversive and therefore motivate problem-solving.

An equal (and long) amount of time was expended by participants on incongruent valid trials regardless of whether the belief-content conflict cue was embedded or not, or whether this cue appeared as the major or minor premise. Consequently, it seems that the location of the belief-content conflict cue does not boost reasoning in the sense of encouraging more System 2 activity, but rather that the cue may inspire a different application of System 2 resources, for instance by encouraging forwards as opposed to backwards reasoning, or complex multidimensional thinking as opposed to binary relational thinking.
The pressure that may have led the belief-content conflict cue to be most effective as a minor premise in valid syllogisms was not a pressure that participants faced when evaluating invalid syllogisms. To reiterate, participants may have engaged in more sophisticated System 2 reasoning when belief-content conflict was presented in close rather than far proximity to intuitively sensed belief-logic conflict in valid arguments. This may have been because the conflict was concentrated rather than dispersed across the argument, therefore putting more pressure for resolution on to the reasoner. In contrast, the invalid trials containing belief-content conflict would not have pressured participants to the same degree because they were congruent and participants would have been intuitively aware of this accordance between content and structure. Participants would therefore have been likely to go with their beliefs and yield a content-driven heuristic judgment. It would presumably require an early cue to disrupt a reasoner’s tendency to apply a content-oriented System 1 process. If the cue occurs as the minor premise in the argument it may be too late for the reasoner to switch strategies, for instance by trading a content-oriented System 1 process for a structurally-oriented one such as matching or the min-heuristic. Based on a review of the reaction time measures, it seems reasonable to conclude that the belief-content conflict cue, when embedded in invalid and congruent trials, disrupted the status quo and inspired reasoners to switch gears within System 1 as opposed to initiating a System 2 strategy. Specifically, no increase in reaction times was observed between congruent trials with or without the belief-content conflict cue.
CHAPTER 7:
BELIEF-CONTENT CONFLICT IMPACTS SEMANTIC PROCESSING

Study 1B

Rationale and Expectations

Study 1B follows up on questions raised by Study 1A relating to participants’ perceptions of the plausibility/implausibility of aggregate syllogisms. Pre-testing focused on participants’ perceptions of the plausibility/implausibility of individual propositions. More specifically, during pre-testing participants were asked to evaluate propositions outside of the syllogistic framework, a task that mimicked the initial conditions under which current participants viewed the major and minor premises. Pre-testing indicated that on average the plausible propositions in Study 1A should have been perceived as highly plausible and unbelievable propositions should have been perceived as highly implausible. The question that Study 1B addresses is whether or not plausibility ratings of individual propositions still apply when those propositions are in mixed company (i.e., in the company of neighboring premises within a three part syllogism). Specifically, when a proposition containing belief-content conflict is included in a syllogism, is this statement, now in the company of other propositions, still perceived as implausible? Or, is this statement perhaps eclipsed by the other parts of the argument? For instance, do participants perhaps skip over the belief-content conflict when it is partnered with an implausible conclusion?

It is of specific importance to explore the role of belief-content conflict within valid syllogisms that begin with either a universal or particular statement. Belief-content conflict in the major premise of valid syllogisms was associated with significantly lower accuracy than when it appeared as the minor premise. The question that needs to be answered here is whether participants noticed the belief-content conflict and implausible conclusion equally in these two
conditions (UBU and BUU). If this was not the case then we might conclude that a warning signal’s proximity to the conclusion (far UBU vs. close BUU) determines its efficacy as a cue. A warning signal immediately preceding an implausible conclusion (BUU) may be more successful as a cue than one separated from the conclusion by plausible content (UBU), possibly because in the former case the cue diminishes participants’ awareness of the conclusion thereby decreasing belief bias.

In the case of invalid syllogisms beginning with a particular statement, accuracy was found to be no different between the congruent (BBU) and warning signal (BUU) conditions. Belief-content conflict should have lowered accuracy in this latter condition if it led participants to abandon heuristic, belief-oriented processing that predominated in the congruent condition. This was not the case which suggests that the warning signal performed poorly as a cue when presented as a minor premise, or alternately that participants were less at the mercy of syllogistic content when they were evaluating congruent syllogisms beginning with a particular statement. The key question then is whether or not the belief-content was noticed by participants in the Invalid x Particular x warning signal (BUU) condition. If the belief-content conflict was not noticed then it seems reasonable to suppose that participants processed the congruent and warning signal conditions similarly which could explain the null difference between their respective mean accuracy scores on these two conditions. If however the warning signal was noticed, then there was a content difference. The interpretation I would then provide for the Believability x Major Premise interaction on accuracy for invalid syllogisms would be that although participants were aware of belief-content conflict when presented in the major and minor premises, they were not induced to replace default, content-oriented heuristic processing with structure-oriented heuristic processing unless the cue appeared as the major premise.
The second part of Study 1B attempted to determine whether the presence and location of belief-content conflict in syllogisms influenced participants’ preferences for reviewing (a) evidence substantiating syllogistic content, and (b) Venn diagrams elucidating syllogistic structure. In Study 1A it was found that belief bias was attenuated by belief-content conflict in the premises, possibly because the conflict served as a warning signal that prompted participants to adopt a structure-oriented, System 2 driven reasoning strategy. In addition to elevated rates of accuracy for incongruent valid arguments embedded with a belief-content conflict cue, long reaction times were also observed in such trials, a set of observations that would indicate working memory load and System 2 activity. It is expected that participants might prefer to view structural diagrams in this condition relative to conditions in which belief bias was not attenuated. When belief-content conflict was introduced into invalid, congruent trials it succeeded in attenuating belief bias as the major premise, yet reaction times were short. This combination of observations would implicate System 1 and imply that heuristic processing was applied to either content or structure of the problem. Handley and colleagues (2011) emphasized that although System 1 is usually linked with belief- and content-oriented processing that there is evidence to suggest that Systems 1 and 2 perform “fast” and “slow” surveys of content and structural problem features, respectively. Thus, participants’ preferences for evidence versus Venn diagrams may shed light on the mechanisms underlying their judgments of invalid problems with and without belief-content conflict.
Methods

Participants

The sixty-two undergraduate students (26 males, 36 females) who participated in Study 1A also took part in Study 1B. Their involvement was remunerated by way of a .5 credit towards their final course grade in Introductory Psychology.

Materials and Measures

Eight test syllogisms (4 valid, 4 invalid) from Study 1A were used. In terms of beliefs – congruent, incongruent, and warning signal cases were compared for valid and invalid syllogisms (BBB, BBU, UBU/BUU; BBU; BBB; UBU/BUU, respectively). Participants were asked to identify implausible parts of the syllogisms. They also completed a forced choice task during which they indicated whether they would prefer to see evidence to substantiate the syllogistic content or a Venn diagram to illustrate the syllogistic structure.

Procedure

PowerPoint was used to coordinate Study 1B. Participants were instructed to read the material on each slide before advancing to the next slide by clicking on it in the slide browser located on the left side of the screen. Participants reported familiarity with PowerPoint and were observed to maneuver their way through the introductory slides without difficulty. Participants were informed about the difference between syllogistic content and structure. They were also told that one can “check the facts” or seek out information (e.g., by consulting an encyclopedia) to verify syllogistic content and that syllogistic structure can be verified by consulting Venn diagrams and other visual models (see Appendix F).

After reviewing this information participants were presented with individual syllogisms. Participants were informed as to the validity or invalidity of each syllogism. They were then
asked to indicate whether they believed or did not believe the syllogistic content based on their everyday knowledge about the nouns/adjectives within the syllogism (e.g., boats, ships, and canoes). In cases where the syllogisms were judged to contain implausible material, participants were prompted to identify the specific parts of the argument that they did not believe by typing into the Notes section of the PowerPoint display. Finally participants indicated whether they preferred to see evidence substantiating content or a Venn diagram illustrating structure.

Design

To prevent participants from becoming fatigued, this follow-up study contrasted a subset of trial types rather than presenting all possible combinations arising from an interaction between Validity, Believability, and Major Premise. The key trial types that were contrasted in Study 1B attempted to elaborate on the key findings reported in Study 1A by comparing valid incongruent trials in which belief-content conflict appeared as the major (UBU) or minor (BUU) premise, and invalid congruent trials opening with particular statements in which belief-content conflict was either present (BUU) or absent (BBU).

Results

The Warning Signal in Valid Syllogisms

Pre-testing determined that when participants evaluated individual implausible propositions, they were perceived to be equally unlikely. In order to determine whether or not participants’ perceptions of such propositions changed once these statements were embedded into arguments, the frequency with which participants identified the conclusion and warning signal propositions as implausible was tabulated for valid syllogisms of the forms UBU and BUU. This follow-up task differed from pre-testing because participants did not rate the plausibility of the propositions, they merely identified the parts of the argument that they did not
believe. A paired-samples t-test revealed that there was a significant difference between participants’ detection of implausibility in arguments depending on the location of the belief-content conflict. Participants were approximately 20% more likely to notice the conclusion’s implausibility when belief-content conflict appeared in the major premise (M= .72, SD=.45) compared to the minor premise (M=.51, SD=.51); t(60) = 2.74, p<.01 (see Figure 11).

Figure 11

*Implausible Conclusion Perceived as Unlikely More Often when Warning Signal Comes at Outset of Argument in Valid Syllogisms*

Participants were approximately 20% more likely to note the belief-content conflict cue itself when it preceded the conclusion as the minor premise (M=.57, SD=.50) compared to being embedded in the major premise (M=.38, SD=.49); t(60) = -2.83, p<.01 (see Figure 12).
Interpretations about the Warning Signal in Valid Syllogisms

Although syllogisms may be constructed out of comparable parts, it seems as though the clauses interact with each other when participants consider the syllogisms holistically. In valid syllogisms containing an implausible conclusion preceded by belief-content conflict, participants’ sensitivity to the impossibility of both the conclusion and the belief-content conflict depended on the location of this cue. When an argument contained belief-content conflict in its major premise, participants’ accuracy when judging its structure was moderate, approaching 60% (as per Study 1A), they successfully identified the conclusion’s implausibility about 72% of
the time, and were relatively poor at identifying that the warning signal itself was implausible. In contrast, when an argument started out with a neutral major premise followed by two implausible statements – belief-content conflict in the minor premise and the conclusion – participants had high accuracy when judging the argument’s structure (approximately 80% as per Study 1A), were often unsuccessful at identifying that the conclusion was implausible, and were more likely to detect implausible content inherent within the belief-content conflict.

Taken together, these findings suggest that belief-content conflict is a salient problem feature and that when it appears in close proximity to an implausible conclusion it detracts participants’ attention from that conclusion’s content. One might extend this interpretation by speculating that belief-content conflict is a problem of categorization, one that poses an analytic challenge to participants when present within a syllogism. Participants might have time to resolve this categorization problem before arriving at the conclusion when this conflict appears as the major premise. In this way they would be able to devote ample cognitive resources towards content-oriented processing of the conclusion. Comparatively, participants might not have time to resolve the categorization problem before arriving at the conclusion when this conflict directly precedes the conclusion. In this case, upon encountering the conclusion content, participants might still be working on the belief-content conflict and have fewer resources to put towards content-oriented processing of the conclusion. If this were the case, it would be expected that participants would be less vulnerable to the biasing effects of conclusion content because they were giving less thought to it in the first place.

*The Warning Signal in Invalid Syllogisms*

Pre-testing established that the implausible conclusions that were used to build the invalid syllogisms in Study 1A were on par in terms of perceived plausibility when each statement was
evaluated on its own. Whether or not the conclusions were identified as implausible with comparable frequency upon being embedded into invalid syllogisms of the forms BBU (congruent) and UBU/BUU (warning signal) was tested through paired-samples t-testing. These specific comparisons were made because participants’ judgment of accuracy differed significantly between the congruent and warning signal conditions for syllogisms opening with universal statements but not for particular statements. This raised the issue of whether the warning signal in the BUU condition was ineffective at moderating belief bias, or whether arguments beginning with particular statements were poor at inducing the belief bias effect and therefore were a poor comparison condition.

T-testing revealed that for invalid syllogisms starting with a universal statement, participants were much more likely to acknowledge the conclusion’s implausibility in the BBU (M=.90, SD=.28) as compared to the UBU condition (M=.67, SD=.48); t(60) = 4.09, p<.001. When the invalid syllogism began with a particular statement, there was no difference between the BBU and warning signal condition, BUU – participants were equally likely to notice that the conclusion was implausible, 56% and 62% of the time, respectively, t(60) = -.85, p=.40 (see Figure 13). Finally, there was a significant difference between the perceived plausibility of conclusions in the BBU condition depending on whether the argument started with a universal or particular statement, t(60) = -4.96, p<.001. Participants identified implausible conclusions the most frequently when they were embedded in arguments beginning with universal statements as compared to particular statements.
**Interpretations about the Warning Signal in Invalid Syllogisms**

The finding that participants were less aware of implausible conclusion content in the UBU condition relative to the congruent condition (BBU) suggests that belief-content conflict functions by diminishing content-oriented processing of the conclusion. The BUU condition did not attenuate belief bias relative to the BBU condition for trials starting with particular statements, nor was this warning signal condition associated with reduced sensitivity to conclusion content. This finding supports the interpretation that belief-content conflict, when it is a successful cue, counteracts belief bias by discouraging content-oriented processing. In the case of BUU where the cue did not succeed, belief bias was not attenuated nor was there diminished
awareness of conclusion content. Finally, the finding that universal propositions in arguments are associated with greater sensitivity to conclusion content compared to particular major premises contributes to our understanding of the null effect of belief-content conflict as the minor premise in Study 1A. It is true that this trial type failed to reduce conclusion content-orientation, yet part of the reason that the minor premise’s belief-content conflict may appear to be an ineffective warning signal is because it was embedded within an argument that opened with a particular statement. Given that arguments opening with universal statements were more likely to promote sensitivity to conclusion content that might induce belief bias, it seems that arguments starting with particular statements were by their very nature less likely to lead to a bias that would go on to be attenuated by a warning signal. Thus, the problem with minor premise belief-content conflict in invalid trials was two-fold: the structure of the argument was less likely to press belief-content conflict into action as an attenuator of belief bias and in cases where this cue was required (belief bias was in evidence), it was poor at diminishing participants’ orientation towards conclusion content.

Evidence versus Venn Diagram Preferences

Validity (valid, invalid) and Believability (BBB, BBU, UBU/BUU) were entered into an ANOVA model as within-subjects variables where preference (for evidence or a Venn diagram) was the dependent measure. There was no significant preference difference between valid and invalid syllogisms. In both cases participants were more interested in viewing evidence to substantiate implausible rather than plausible semantic propositions, $F(2, 116) = 23.69, p < .001, \eta^2 = 0.16$ (see Figure 14). Regardless of whether or not the syllogism contained belief-content conflict prior to the implausible conclusion, participants wanted to see evidence approximately 67% of the time as compared to when the syllogism was wholly believable, in which case they
were only interested in reviewing evidence about 30% of the time. Given that the preference judgment was a forced-choice task, the previous finding corresponds with the finding that participants wanted to see a Venn diagram 70% of the time after they had encountered a wholly plausible syllogism, yet only 33% of the time following their review of syllogisms containing implausible content.

Figure 14

*Main Effect of Believability on Preference for Evidence*

*Summary of Findings and Preliminary Interpretations*

In valid syllogisms, belief-content conflict serves as effective warning signal when presented as the minor premise, directly preceding the implausible conclusion. In this position, the implausibility of the warning signal may be more salient than if it were at the opening of the argument, and it may lead participants to focus less critically on the conclusion because they are
in the midst of resolving the categorical challenge posed by the minor premise. Invalid syllogisms that contained an implausible conclusion statement were more likely to have their conclusions identified as implausible when they started with a universal statement as compared to a particular statement. It appears that belief-content conflict attenuates belief bias by diminishing participants’ sensitivity to conclusion content, a role that is effectively enacted when belief-content conflict occurs as the major premise but not as the minor premise. Furthermore, placing belief-content conflict within the minor premise of the invalid syllogisms in this study may have disadvantaged an already weakened cue by casting it into a situation in which it was less likely to be called upon to counteract belief bias.

Although belief-content conflict was shown to diminish the salience of implausible content in the conclusion under certain conditions, no evidence emerged from the forced-choice task to substantiate our hypothesis that it encourages a focus on structure and simultaneously discourages a content-based reasoning strategy. The warning signal conditions were not associated with a greater proclivity by participants to review Venn diagrams as compared to the BBU conditions for valid and invalid syllogisms. The key difference that the forced-choice task brought to light was that implausible content, regardless of where it falls in the syllogism or how many parts of the syllogism it fills, urges participants to prefer evidence. On the other hand, participants preferred Venn diagrams more than evidence after reviewing syllogisms that were wholly plausible.

The forced-choice task may not provide evidence to support my hypothesis that warning signals promote a structural orientation during reasoning because this task did not take place during the syllogistic reasoning task. As such, what participants are interested in after they have solved a syllogism may not be the same as what they are focusing on while solving the
syllogism. Faced with implausible content in a syllogism’s major or minor premise and its conclusion, participants may avoid focusing on their beliefs while solving the syllogism even though the implausible content is registered. The warning signal and conclusion might be recognized as implausible, but participants might consciously avoid coasting along heuristically and instead focus explicitly on structural evaluations of the syllogism’s premises and conclusion. Participants might have surmounted their beliefs during the reasoning task, only to let their curiosity about unresolved semantic content take over during the post-reasoning task when they are asked what they would prefer to learn more about.

One might wonder then, why the BBU condition is no different in terms of evidence preference from the warning signal conditions. The BBU conditions are the traditional belief bias conditions where participants are known to rely on content as a heuristic. Presumably, content should not be interesting to participants who have had ample time to think about it given that it was central to their reasoning process. In this instance, participants may prefer to see content, not because it will reduce uncertainty about the semantic content of propositions that had been shelved to make way for a structurally-guided assessment of the syllogism, but rather, because evidence may be a way of double-checking whether one’s judgment about validity was accurate. I would conclude that a preference to see evidence may be the product of two motivations – a desire to learn about content that was ignored during an earlier structurally-oriented task, or a desire to verify the accuracy of one’s earlier, semantically-guided, judgment. These two motivations are consistent with the distinction between learning and performance orientations in the problem-solving literature (Eliot & Dweck, 1988; Vansteenskiste, Lens, & Deci, 2006).

An alternate explanation for why the forced-choice task yielded no preference for Venn diagrams in the warning signal conditions is that the task-set itself may have discouraged
participants from being curious about structure, regardless of condition. Given that participants were explicitly informed as to the structure (valid, invalid) of each syllogism, they may have had no interest in seeing this structure depicted as a Venn diagram. It is possible to counter this explanation – people often know the answer to problems and even so find them motivating to think about. As such, it is possible that participants could know that a structure is valid or invalid and still want to see the Venn diagram.

Based on my own reflection, it seems that problems come in two varieties – one in which the product of reasoning is emphasized, and another in which the process of reasoning is emphasized. I would say that semantic tasks emphasize the product – for instance a memory or declarative knowledge task emphasizes the answer. The answer is a fact or piece of information that one does or does not know; is or is not able to bring to mind. Similarly, the semantic side of syllogistic reasoning emphasizes a categorical judgment whereby the reasoner decides whether X does or does not belong to Y; the answer is yes or no. In these instances, knowing the answer does not motivate the problem-solver to solve the problem (e.g., solving a crossword puzzle is no fun if one has reviewed the solution already). On the other hand, if the task is abstract and emphasizes the process rather than the product, then I would say that knowing the product may not diminish said process (e.g., if one is solving a math problem, knowing the answer is often something that facilitates the process without entirely diminishing the difficulty of the task at hand; if one is making an origami bird, knowing what the end result looks like can promote the process). These observations lead me to suppose that participants might be interested in seeing the Venn diagram even when they know that the structure is valid/invalid, because the Venn diagram elucidates the process of problem-solving rather than merely reiterating the product or structural judgment that they encountered a priori.
CHAPTER 8: EFFICACY OF BELIEF-CONTENT CUE
DEPENDS ON SYLLOGISTIC DIFFICULTY LEVEL

Study 2

Rationale and Expectations

Study 1A explored valid and invalid syllogisms of the forms BBB, BBU, and UBU/BUU and showed that under some conditions belief-content conflict appears to attenuate belief bias. Study 1B established that belief-content conflict may function as a cue by diminishing participants’ sensitivity to conclusion implausibility. Although it seems plausible to suppose that the belief-content conflict cue may encourage structurally-oriented processing by System 2, this was not detected by Study 1B. The effect of belief-content conflict when present in the major premise of invalid congruent syllogisms led to lower accuracy relative to strictly incongruent invalid trials. This occurred because the belief bias that would normally artificially inflate accuracy in the congruent condition was diminished.

In light of these findings, the purpose of Study 2 was firstly, to explore invalid syllogisms of the forms BBU, BBB, and UBB. This design allowed us to answer the question of whether or not belief-content conflict attenuates belief bias, thereby promoting accuracy when participants evaluate invalid syllogisms with plausible conclusions. A secondary purpose of Study 2 was to introduce “difficulty level” as an independent variable that may impact participants’ performance during syllogistic reasoning tasks in which there was variation in terms of the conclusion’s plausibility as well as the presence or absence of belief-content conflict in the premises of the argument. A syllogism’s structure can be more or less difficult for a reasoner to judge; syllogisms that contain negative propositions (e.g., No A are B, Some A are not B) have been shown to be significantly more difficult to judge compared to those containing affirmative
propositions (Johnson-Laird, 1983). Difficulty level has also been implicated as a key factor in reasoning outcomes by Feather (1964) who showed that belief bias is high when syllogisms are challenging given one’s poor logical skills or because they entail inherent structural complexities (Markovits & Nantel, 1989). Oakhill, Johnson-Laird and Garnham (1989) addressed the possible relationship between beliefs and difficulty level, suggesting that content may urge participants to engage in logical reasoning but that excessive difficulty may block this intention. In other words, the will may be present but the way may be absent.

It is expected that invalid syllogisms with a plausible conclusion of the form BBB will be resolved with poor accuracy because beliefs will guide participants to falsely judge the argument as being true because the conclusion is believable. Furthermore, it is expected that upon introducing belief-content conflict in the major premise, a boost in accuracy will be observed for invalid syllogisms of the form UBB because the implausible major premise will prompt participants to be more vigilant and less reliant on a semantically guided reasoning strategy when evaluating the argument. In terms of the difficulty level of the syllogisms, it is expected that a semantically guided reasoning strategy will predominate for difficult syllogisms where structural consideration of the argument is very labour-intensive. This would mean that beliefs may be better able to bias structural judgments for difficult syllogisms, whereas participants may be less vulnerable to syllogistic content, and better able to wield a non-semantic strategy when the syllogism’s structure is easier to resolve.

Methods

Participants

Sixty-six undergraduate students (31 males; 35 females) enrolled in Introductory Psychology at UTSC took part in this study in exchange for a .5 credit towards their final course
grade. Participants from Studies 1 and 2 were restricted from the current study. All participants in the current study were native English speakers or had begun speaking English as a second language by the time they were 3 years of age. No participants had previously received training in logic or syllogistic reasoning.

Materials

Two non-semantic syllogisms (one valid; one invalid), were presented as warm-up trials; and six semantic filler trials were interspersed between 24 semantic test trials. The test trials varied in structure (valid; invalid), difficulty level (easy; difficult), and congruency of the conclusion (congruent; incongruent without a warning signal; incongruent with a warning signal). Valid syllogisms were therefore embedded with Believable (B) and Unbelievable (U) propositions according to the following pattern: BBB; BBU; and UBU, whereas invalid syllogisms were embedded with content in one of three ways: BBU; BBB; and UBB. These conditions allowed an explicit test of the following hypotheses: (a) that congruency inflates accuracy and that incongruency diminishes accuracy, and (b) that a warning signal may attenuate belief bias thereby boosting accuracy if one were to compare the BBU and UBU conditions for valid syllogisms or the BBB and UBB conditions for invalid syllogisms. Difficulty level was determined based on pre-testing which confirmed that syllogisms containing a negative proposition (No A are B) are considerably more difficult than affirmative ones. Based on this design there were 12 possible combinations, each of which was represented by two different syllogisms that were administered within-subjects (see Appendix G for Study 2 design and Appendix H for the content of the test syllogisms).
Measures

Of the practice and filler trials, four of the syllogisms were of an ‘easy’ level (i.e., 90% of participants during pre-testing solved these syllogisms accurately). An accuracy score per ‘easy’ trial was generated for participants in the current study. Participants with below-average aptitude for syllogistic reasoning were identified by referring to individual accuracy for these trials and in the end one male participant failed to solve all four of these easy syllogisms leading to the exclusion of his data from further analyses.

For each trial participants indicated whether the conclusion was a necessary product of the premises. Accuracy (0 = miss; 1 = hit) for participants’ validity judgments, the amount of time (msec) that participants spent solving the syllogism, and their self-reported confidence about their evaluation of the argument on a scale of 1 to 5 (1 = not at all confident; 5 = extremely confident) were the primary dependent measures in this study. Participants were also asked to evaluate the believability of the 72 propositions making up the test syllogisms on a scale of 1 to 7 (1 = highly implausible; 7 = highly plausible).

Procedure

Participants were seated at individual computer terminals and were asked to turn off their cellular phones. They were introduced to the syllogistic task through written instructions that appeared on the computer screen. Participants were encouraged to take their time reading the instructions and to advance through the instructions at will by pressing the space bar. The trials were programmed using E-Prime. At the beginning of the session participants were introduced to the logical syllogism according to the same protocol used in Study 1A (Appendix E). Practice trials were administered before participants proceeded to the test phase of the study during which they evaluated 24 test trials interspersed with filler trials. Again, the trials were administered in
the same way as was outlined in Study 1A, with participants moving at will through the premises before using specified keys on the keyboard to evaluate the syllogism’s structure and to indicate their confidence about this judgment. Participants had unlimited time to complete each trial and the length of time they spent judging the overall argument was logged by E-Prime. A second test trial block was administered to participants. In this case they were asked to evaluate the believability of individual propositions rather than judging the structure of whole arguments. The keyboard’s number pad (from 1 to 7) was employed by participants to log their responses.

Design

Gender was a between-subjects variable and the within-subjects variables, administered with repeated measures, were structure (valid; invalid), difficulty level (easy; difficult); and congruency of the conclusion (congruent; incongruent without a warning signal; incongruent with a warning signal). The test syllogisms and test propositions represented two blocks of trials. For Order 1 the items within each block were each randomly sequenced and the syllogism trials were administered prior to the proposition trials. For Order 2 the sequence within each block was reversed and the syllogism trials were again presented prior to the proposition trials. Orders 3 and 4 saw the propositions presented prior to the syllogisms; the sequence of items within the blocks was identical to that in Order 1 and Order 2, respectively.

Results

Manipulation Checks

Participants in this study evaluated the believability of the individual propositions making up the test syllogisms on a 7-point rating scale. The average plausibility of a believable (B) proposition was 90% whereas unbelievable (U) propositions were seen as plausible only 32% of the time. This difference was significant ($p<.0001$) and consistent across the conditions.
of the study. In other words, the average plausibility of the individual propositions comprising the test syllogisms was comparable ($p > .05$) across conditions (validity, difficulty, congruency).

The average plausibility of each syllogism in the study was calculated by adding up the participants’ ratings per proposition. Each proposition could have received a total rating of 7 points if it was deemed to be highly plausible and each proposition was evaluated 66 times. The overall plausibility of a single syllogism was therefore calculated by adding up the 66 ratings of the syllogism’s 3 propositions and dividing that number by the total possible points (1386) that the argument could have obtained if it were judged to be 100% plausible by all the raters. The equation is: \( \frac{\text{Sum of ratings from Major + Minor + Conclusion}}{1386} \times 100\% \). Consistent with the manipulation in this study, the plausibility of the arguments differed systematically according to semantic structures used in this study (BBB, BBU, UBB, UBU). Arguments with three believable propositions (BBB) had an average plausibility rating of 91%. Those with two believable propositions (BBU and UBB) were judged to be 67% and 69% plausible, respectively ($p > .05$). Finally, arguments with only one plausible proposition (UBU) were rated as 48% plausible. Perceived plausibility differed significantly for arguments containing three plausible statements, two plausible statements, and one plausible statement, $p < .05$; and, each argument type was found to be semantically distinct in terms of this perceived quality.

*The Valid Syllogisms*

A repeated measures analysis of variance was employed to determine if there were any significant main effects of or interactions between the independent variables which were Gender, Difficulty, and Congruency on participants’ accuracy, time spent evaluating the syllogisms, and self-reported confidence. Tables 9, 10, and 11 list the main effects of these variables on valid trials. Appendix I summarizes non-significant results related to Difficulty and Congruency.
Table 9

*Main Effects of Difficulty on Accuracy, Reaction Time, and Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficult</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.63</td>
<td>.47</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>14041.44</td>
<td>18875.04</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.24</td>
<td>4.03</td>
<td>.01</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 10

*Main Effects of Congruency on Accuracy, Reaction Time and Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Congruency</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.82</td>
<td>.14</td>
<td>.68</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>14707.83</td>
<td>18078.10</td>
<td>16588.69</td>
<td>.05</td>
<td>*</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.31</td>
<td>3.95</td>
<td>4.14</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Mauchly’s test of sphericity showed that the variances were significantly different ($p < .05$). The Greenhouse-Geisser degrees of freedom and p-value were referred to.

Table 11

*Main Effect of Gender on Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>4.39</td>
<td>3.88</td>
<td>.01</td>
<td>ns</td>
</tr>
</tbody>
</table>
Accuracy. The difficulty level, $F(1, 63) = 35.19, p < .001, \eta^2 = .03$, and congruency of the conclusion, $F(2, 126) = 172.10, p < .001, \eta^2 = .43$ (see Figure 15) in valid syllogisms significantly impacted participants’ accuracy when judging whether the conclusion necessarily followed from the premises of the argument. Difficult syllogisms were accurately solved only 47% of the time whereas easier syllogisms, lacking a negative proposition (no A are B), were solved correctly by participants about 63% of the time. Post hoc testing for the three-level Congruency variable revealed that the BBB, BBU, and UBU conditions were distinctly different from each other, being solved accurately 82%, 14%, and 68% of the time, respectively. This main effect replicates the effect of Believability on accuracy that was obtained during Study 1A.

Figure 15

Main Effect of Congruency on Accuracy for Valid Syllogisms
An interaction between the difficulty of a valid argument and the congruency of its conclusion was also significant, $F(2, 126) = 14.21, \eta^2 = .03$ (see Figure 16). Post hoc testing revealed that participants were (a) most accurate when solving congruent syllogisms (BBB) that were easy rather than difficult; (b) equally poor when solving incongruent syllogisms (BBU), regardless of the difficulty level of the syllogism, and (c) that the warning signal condition (UBU) significantly improved accuracy relative to the incongruent condition for both difficulty levels, but that this effect was greatest for the easy syllogism type. No other main effects or interactions were significant.

Figure 16

*Interaction Effect of Difficulty x Congruency on Accuracy for Valid Syllogisms*

*Reaction time.* The amount of time participants spent judging whether the conclusion followed necessarily from the premises differed substantially depending on the difficulty level of the argument, $F(1, 64) = 18.20, p<.001, \eta^2 = .03$, the congruency of its conclusion, $F(2, 128) =$
4.41, $p<.05$, $\eta^2 = .01$ (see Figure 17), and the interaction between difficulty and congruency, $F(2, 128) = 8.00, p<.01, \eta^2 = .02$ (see Figure 18).

Figure 17

*Main Effect of Congruency on Reaction Time for Valid Syllogisms*

Participants spent significantly more time evaluating difficult syllogisms compared to easy syllogisms. Post hoc testing for the main effect of congruency revealed that participants spent the least amount of time evaluating the congruent syllogisms and significantly more time on the incongruent syllogisms regardless of whether or not a warning signal was present. Post hoc testing for the interaction effect revealed that for easy syllogisms participants spent significantly more time on incongruent trials, regardless of the presence or absence of belief-content conflict in the premises, and less time on the congruent trials. In contrast, for difficult syllogisms the
belief-content conflict in the premises was associated with significantly shorter reaction times compared to the congruent and incongruent (non-warning signal) conditions.

Figure 18

*Interaction Effect of Difficulty x Congruency on Reaction Time for Valid Syllogisms*

![Graph showing reaction time by difficulty and congruency levels](image)

Confidence. Main effects of Difficulty $F(1, 63) = 12.03, p<.01, \eta^2 = .02$, Congruency $F(2, 126) = 10.50, p<.001, \eta^2 = .03$ (see Figure 19), and Gender $F(1, 63) = 13.47, p<.01, \eta^2 = .004$, were obtained for participants’ self-reported confidence about their validity judgments. Participants were significantly more confident when evaluating ‘easy’ syllogisms compared to ‘difficult’ syllogisms. Male participants were significantly more confident than female participants that they had accurately judged the validity of the test syllogisms. Finally, post hoc testing revealed that congruent syllogisms (BBB) were associated with the highest level of confidence and incongruent syllogisms without a warning signal (BBU) were associated with substantially lower levels of confidence. Interestingly, when an incongruent conclusion statement was preceded by belief-content conflict (UBU), self-reported confidence was significantly
greater than when this conflict was absent. The confidence associated with evaluations of UBU syllogisms was not on par with that of BBB evaluations however.

Figure 19

*Main Effect of Congruency on Confidence for Valid Syllogisms*

A repeated measures analysis of variance was applied to test for significant main effects of and interactions between the independent variables, Gender, Difficulty and Congruency, on participants’ accuracy, response time, and self-reported confidence. Main effects are summarized in Tables 12, 13 and 14. No interaction between Difficulty and Congruency was found (see Appendix I).
### Table 12

**Main Effects of Difficulty on Accuracy, Reaction Time, and Confidence for Invalid Syllogisms**

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficult</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.58</td>
<td>.46</td>
<td>.01</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>20819.66</td>
<td>14531.78</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.14</td>
<td>3.91</td>
<td>.01</td>
<td>ns</td>
</tr>
</tbody>
</table>

### Table 13

**Main Effects of Congruency on Accuracy, Reaction Time, and Confidence for Invalid Syllogisms**

<table>
<thead>
<tr>
<th>Congruency</th>
<th>BBU</th>
<th>BBB</th>
<th>UBB</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.77</td>
<td>.28</td>
<td>.51</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>17186.70</td>
<td>15767.91</td>
<td>20072.56</td>
<td>.05</td>
<td>*</td>
</tr>
<tr>
<td>Confidence</td>
<td>3.99</td>
<td>4.17</td>
<td>3.92</td>
<td>.05</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Mauchly’s test of sphericity showed that the variances were significantly different ($p<.05$).

### Table 14

**Main Effects of Gender on Reaction Time and Confidence for Invalid Syllogisms**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Time</td>
<td>20268.34</td>
<td>15083.12</td>
<td>.05</td>
<td>ns</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.30</td>
<td>3.75</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>
**Accuracy.** Main effects of Difficulty $F(1, 63) = 12.10, p < .01, \eta^2 = .02$ and Congruency $F(2, 126) = 59.64, p < .001, \eta^2 = .19$ (see Figure 20) were obtained for participants’ accuracy when evaluating invalid syllogisms to determine whether an argument’s conclusion followed necessarily from its premises. Participants were more successful when evaluating ‘easy’ as opposed to ‘difficult’ syllogisms. Post hoc testing revealed that accuracy approached 80% for congruent trials (BBU), was substantially lower for incongruent trials (BBB) at 28%, and was significantly improved, yet at chance levels (51%) for the warning signal trials (UBB).

Figure 20

*Main Effect of Congruency on Accuracy for Invalid Syllogisms*

*Reaction time.* The amount of time that participants spent reasoning about whether the conclusion of an argument followed necessarily from its premises varied significantly depending
on Gender, $F(1, 63) = 4.22, p<.05, \eta^2 = .02$, Difficulty of the argument, $F(1, 63) = 35.68, p<.001, \eta^2 = .06$, and the Congruency of the conclusion, $F(2, 126) = 4.63, p<.05, \eta^2 = .02$ (see Figure 21). Specifically, male participants had longer reaction times than female participants and participants spent more time thinking about easy syllogisms than about difficult ones. Post hoc testing revealed that participants spent significantly more time evaluating incongruent syllogisms containing belief-content conflict in the premises than they did when reviewing incongruent syllogisms without this warning signal, or congruent syllogisms.

Figure 21

*Main Effect of Congruency on Reaction Time for Invalid Syllogisms*

Confidence. Participants’ self-reported confidence about their judgments of invalid syllogistic structure varied significantly according to Gender, $F(1, 63) = 17.89, p<.001, \eta^2 = .005$, the Difficulty level of the argument, $F(1, 63) = 11.67, p<.01, \eta^2 = .02$, and the Congruency
of the conclusion relative to the structure of the argument, $F(2, 126) = 4.23, p < .05, \eta^2 = .02$ (see Figure 22). Male participants were significantly more confident than their female counterparts. Participants reported more confidence about their judgments of easy as opposed to difficult syllogisms. Post hoc testing revealed that participants expressed the most confidence in the incongruent condition (BBB) relative to both the congruent (BBU) and incongruent warning signal (UBB) conditions which did not differ from each other.

Figure 22

*Main Effect of Congruency on Confidence for Invalid Syllogisms*

### Summary of Findings and Preliminary Interpretations

*The valid syllogisms.* In terms of congruency, valid arguments with implausible conclusions (BBU) were solved less accurately (14%) and less confidently than when they contained belief-content conflict prior to the implausible conclusion (UBU). In this latter
condition, accuracy (68%) and self-reported confidence were substantially higher. This pattern was not observed for the reaction time measure in that incongruent trials, with or without a warning signal, were pondered substantially longer than congruent ones.

The effects of congruency on accuracy and reaction time reported here replicate the effects reported in Study 1A in which the warning signal condition was found to be associated with higher accuracy than the regular incongruent condition and incongruency was affiliated with longer reaction times than congruency. In Study 1A, participants were found to be more confident when judging congruent trials than incongruent trials. Incongruent trials with a warning signal promoted slightly more confidence than those without a warning signal; however, this was not a significant variation. In Study 2, congruent trials were again judged more confidently than incongruent ones, and this time incongruent syllogisms containing a warning signal were judged more confidently than those without a warning signal. This inconsistency might be explained by the fact that in Study 1A the warning signal condition varied in terms of the location of belief-content conflict – for half of the trials it appeared in the major premise and for the other half it was placed in the minor premise and was always a universal affirmative statement. In Study 2 the warning signal was consistently placed in the major premise of the argument and was either a particular affirmative or universal negative statement. This variation in the location, quantity, and quality of the belief-content conflict between Study 1A and Study 2 could account for the emergence of a confidence effect.

Difficulty level of the argument was included as a new variable in Study 2. Valid syllogisms with a negative statement as their major premise (No A are B) were more difficult than syllogisms consisting of affirmative statements only. In this study it was found that difficult valid syllogisms were solved less confidently and less successfully than their easy counterparts
despite the fact that participants took longer on difficult trials. An argument’s difficulty level interacted with the congruency of its conclusion in a way that impacted accuracy and reaction time. Easy syllogisms were puzzled over the longest when they were incongruent, regardless of the presence or absence of the warning signal. This time expenditure didn’t pay off for purely incongruent syllogisms (BBU) where accuracy was substantially lower than in the congruent (BBB) trials. Adding a warning signal into the syllogism (UBU) meant that this input of time paid off more frequently; incongruency preceded by belief-content conflict was resolved more accurately than incongruency without this cue. Difficult syllogisms led to poor accuracy for the incongruent trials compared to the congruent trials, and the presence versus absence of a warning signal didn’t seem to impact task accuracy; BBU and UBU were both poorly judged when the syllogisms were difficult and contained a negative assertion (No A are B); one difference between these two incongruent conditions was that participants pondered the UBU syllogisms for less time compared to the BBU syllogisms.

The difficulty level of an argument in combination with the congruency of an argument’s conclusion and its structure may impact the duration and product of reasoning in a way that differentially favours the presence of a warning signal because a warning signal may only benefit reasoning if it is able to cue participants to employ a non-belief-based reasoning strategy. In the case of easy syllogisms, a warning signal may cue participants into using a non-belief-based strategy – presumably a strategy that relies on a structural assessment of the syllogism. Easy syllogisms are less structurally complex than difficult ones which means that a structural assessment is more apt to result in an accurate validity judgment. Difficult syllogisms have a more complex structure which means that when a warning cue is present and suggests to participants that they should use a non-belief-based reasoning strategy, presumably one that
orients the reasoner to the structure of the argument, this cue is less successful because the underlying structure of the argument is less accessible and more challenging to judge. This finding, one that distinguishes between easy and difficult syllogistic structure, and that suggests a difference in accuracy by virtue of the underlying semantic structure of the syllogism is good evidence for the hypothesis that belief-content conflict cues participants to use a structurally-oriented reasoning strategy to resolve the argument. If this were not the case and participants were not pushed to consider the structure of the argument when faced with the implausible warning signal in the opening of the argument we would not expect to see a difference in the outcome of reasoning about syllogisms that are more and less complex, structure-wise.

The invalid syllogisms. Participants solved invalid syllogisms containing implausible conclusions (BBU) about 80% of the time whereas accuracy was only at 28% for incongruent trials in which an invalid structure led to a plausible conclusion statement (BBB). In both of these instances it is likely that participants’ judgments were heuristically guided by the beliefs inherent in the syllogisms. Participants spent the same amount of time solving these types of syllogisms; reaction time increased substantially for the ‘warning signal’ condition (UBB) and was accompanied by an accuracy rate of 51%. In terms of confidence, the syllogism type containing entirely plausible content (BBB) inspired significantly higher levels of reported confidence than did the trials containing implausible content, regardless of whether this content appeared as the major premise or conclusion. Taken together, it seems as though the BBU and BBB conditions led to clear-cut, albeit heuristic strategies that produced accuracy levels that were significantly above or below chance, respectively. In other words, in these trials participants’ evaluations were the products of systematic divergence from normative reasoning.
as opposed to random momentary lapses in judgment or “mere mistakes” (Stein, 1996, p. 8; Evans, 1984, as cited in Stanovich & West, 2000).

In contrast, when belief-content conflict was embedded in the major premise of invalid and incongruent arguments, it seemed to produce an accuracy rate that was at chance levels, which suggests that participants may have been guessing about the validity of the syllogism. If participants were guessing about the validity of the UBB syllogisms, this is not reflected in the self-reported confidence ratings since confidence associated with the UBB trials was not significantly lower than the other condition containing implausible content (BBU). Following a similar train of thought, participants spent more time on warning signal trials which suggests that if they did end up guessing this was not an entirely thoughtless enterprise which a short RT might have suggested.

An alternative interpretation is that perhaps participants who were faced with belief-content conflict applied different strategies across these ‘warning signal’ trials compared to the strictly congruent and incongruent variety. Perhaps they relied heuristically on content analysis in some of the UBB trials and at other times attempted to assess the underlying structure of the argument. By jumping from one strategy to another rather than employing a single strategy across the trials, participants would end up with an accuracy level on UBB trials that is similar to one that would be expected if they were merely guessing without using any strategy. This accuracy level would not be substantially elevated or reduced compared to chance levels because the exclusive use of a strategy would not polarize the accuracy or lack thereof of their responding.
One way to test this interpretation about strategy-shifting in the UBB trial type would be to explicitly compare the standard deviation for reaction time in the UBB, BBU, and BBB conditions. If a switching strategy were being applied by participants in the UBB trials then we might expect that there would be a greater deviation from the mean reaction time across the data set because some trials would be solved quickly due to guessing and others would be lengthier in duration given a more elaborated, System 2 driven process of analysis. The application of Mauchly’s Test of Sphericity confirmed our expectations in that the standard deviation for RT in belief-content conflict UBB trials was substantially greater than it was for the strictly congruent and incongruent trials. Interestingly, upon comparing easy and difficult trial types this difference in standard deviations was found to apply only for easy trials. This finding supports our earlier suspicion that difficulty level may limit the flexibility with which participants evaluate syllogisms. Easy syllogisms can be assessed heuristically and analytically whereas difficult syllogisms make it less likely that a two-pronged approach to evaluation will be applied; instead ‘guessing’ may be the only possible strategy.

In Study 1A participants were also presented with invalid syllogisms containing belief-content conflict. In that study the warning signal was embedded in an invalid structure that worked up to an implausible statement whereas in Study 2 it was a plausible statement that concluded the argument. The key finding from Study 1A was that the warning signal seemed to attenuate artificially aided accuracy; so, implausible content in the beginning of the argument was associated with greater ‘yea’ saying. Belief bias seemed to be attenuated, as evidenced by the fact that the warning signal condition was associated with reduced accuracy relative to the traditional belief bias condition, BBU, in which conclusion congruency augmented accuracy in a biased manner. ‘Nay’ saying on the other hand, seemed to be more common in Study 2 when a

---

4 As suggested by Dr. Vartanian.
warning signal preceded a plausible statement. In this case, accuracy was substantially greater for the UBB condition than the traditional belief bias condition, BBB, suggesting that belief bias was again attenuated by the warning signal. Study 1B showed that belief-logic conflict may attenuate belief bias by diminishing the salience of the implausible conclusion. Similarly, in Study 2, belief-content conflict may have prompted participants to avoid basing their judgment on the argument’s content, a strategy that seemed to be employed mostly in easy syllogisms where non-heuristic modes of evaluation would be accessible.

One interpretation of the findings from Studies 1A, 1B, and 2 when the warning signal condition is considered for valid and invalid syllogisms is that the warning signal may cue the participant to respond in a way that is opposite to that cued by the semantic meaning of the conclusion statement. A valid syllogism of the form UBU would prompt a “valid” judgment; whereas invalid syllogisms of the forms UBU and UBB would be judged as “valid” and “invalid”, respectively. Although this interpretation would explain the direction of responding observed for participants in both studies, characterized by greater accuracy in the UBU-valid and UBB-invalid trials, and reduced accuracy in the UBU-invalid trials, this interpretation fails to account for the magnitude of the observed change. If participants were using this strategy continuously, accuracy or error would be extreme. For instance, rather than UBB-invalid trials being solved accurately 51% of the time, the reliance on this strategy would be expected to lead to an increase in accuracy that would be markedly above chance levels since the “invalid” judgment would always be accurate for invalid syllogisms of the form UBB. Since the accuracy levels are not polarized in the warning signal condition, it suggests that this simple heuristic is not being applied and instead that a different type of evaluative process is at work.
CHAPTER 9: BELIEF-CONTENT CONFLICT CAN INSPIRE

CONCLUSION-CONTENT BLOCKING

Study 3

Rationale and Expectations

Study 3 attempts to replicate the findings from Study 2 related to participants’ accuracy, reaction time when judging syllogistic structure, and confidence pertaining to this judgment. Study 3 also endeavors to investigate the semantic processing that occurs when participants are evaluating arguments. In particular this study aims to determine how long participants take to judge the everyday plausibility of three-part syllogisms and to test whether the difficulty level of the argument and/or the pattern with which plausible and implausible content is embedded into the argument seems to impact this reaction time. This study also seeks to determine whether there are certain conditions that push participants to focus their attention on the semantics of the conclusion and other conditions that distract participants’ attention from the conclusion and encourage semantic evaluation of other parts of the argument. Study 3 elaborates on Study 1B by (a) presenting participants with a larger sample of arguments to evaluate semantically and (b) measuring reaction times for each semantic evaluation. Study 3 is designed so that we can look at the relationship (and potential trade-offs) between the time spent evaluating the semantics of the conclusion, major, and minor premises of varying types of arguments.

Based on the findings from Study 2, it was hypothesized that in Study 3 belief bias would be observed in the strictly congruent and incongruent trials and that it would be most notably attenuated in easy syllogisms containing belief-content conflict in the major premise. Thus, accuracy should be high in the congruent condition, substantially below chance levels in the incongruent condition, and elevated relative to the incongruent condition when belief-content
conflict serves as an effective warning signal. In terms of semantic evaluation, Study 1B sets up the expectation that belief-content conflict will encourage a de-emphasis of conclusion content. In terms of reaction time for evaluating the plausibility of argument content (as opposed to argument validity), it was expected that participants would take longer to evaluate *highly implausible* arguments (e.g., UBU) compared to *highly plausible* arguments (e.g., BBB). This hypothesis accords with the notion that content that discords with one’s beliefs is highly self-relevant (Hasson, 2005), that self-relevance inspires epistemic vigilance and System 2 activation (Sperber et al., 2010), and finally, that self-relevance diminishes one’s natural, System 1 driven tendency to confirm information (Gilbert et al., 1995).

**Methods**

**Participants**

Fifty-four psychology undergraduate students (25 males, 29 females) from UTSC participated in this study in exchange for .5 credits towards their final grade in Introductory Psychology. Participants from Studies 1A, 1B, and 2 were restricted from the current study. All of the students who participated in this study were native English speakers or had acquired English as a second language by the age of 3 years. No participants had previously been trained in logic or syllogistic reasoning.

**Materials**

The 24 test syllogisms that were employed in Study 2 were used in this follow-up study in combination with 10 filler syllogisms to round out the experimental session. These 10 filler syllogisms represented easy and difficult forms of valid and invalid arguments and were interspersed between the test syllogisms in an effort to disguise the manipulation. The test syllogisms varied systematically in terms of structure (valid; invalid), difficulty level (easy;
difficult), and congruency of the conclusion (congruent; incongruent without a warning signal; incongruent with a warning signal). Valid syllogisms therefore were embedded with Believable (B) and Unbelievable (U) propositions in the following manner: BBB; BBU; and UBU, whereas the pattern in invalid syllogisms was BBU; BBB; and UBB. Given the semantic nature of the 10 additional syllogisms, an effort was made to avoid biasing participants’ responses on the test trials by ensuring that the filler trials were embedded with a mixture of plausible and implausible content. Besides administering the 24 three-part test syllogisms to participants, twenty-four individual propositions – the conclusion statements from the test syllogisms – were excerpted and administered as stimuli in this study.

**Measures**

Participants were instructed to evaluate the real-world plausibility of individual propositions that were either grouped together in three-part syllogisms, or that were presented independently. Specifically, participants indicated whether they believed or did not believe the proposition by pressing the corresponding key on the keyboard, “z” or “m,” respectively. The amount of time spent evaluating the plausibility of each proposition was recorded. Participants also judged the structure of the test and filler syllogisms, indicating whether the conclusion necessarily followed from the premises. The accuracy of this judgment (0 = miss; 1 = hit), amount of time spent solving the syllogism, and self-reported confidence about the structural judgment (1 = not at all confident; 5 = extremely confident) were the primary measures in this study.

**Procedure**

This study was administered using E-Prime and there were three blocks of trials that were presented in a randomized order across the participants in this study. Data were collected
individually. In Block 1 participants judged the structure of 24 test and 10 filler syllogisms and
indicated their confidence about each structural judgment. The amount of time that participants
took to evaluate the structure of the syllogisms was recorded by E-Prime. In Block 2 participants
completed a forced-choice task in which they evaluated the real-world plausibility for each of the
propositions comprising the 24 test syllogisms by logging their judgment using the keyboard
(“z”=plausible; “m”=implausible). Response time for this judgment was recorded by E-Prime.
Block 3 presented participants with individual propositions – the conclusion statements from the
24 test trials. Participants completed the forced-choice task of indicating whether each statement
was plausible (“z”) or implausible (“m”) and response time was logged by E-Prime.

Design

Each Block had two presentation orders that corresponded with one randomized sequence
and its reversal. These orders were presented equally across participants. Additionally, the
sequence of the Blocks’ presentation was counter-balanced across male and female participants
to ensure that all possible orderings of the Blocks were presented. Block 1 investigated the
impact of validity, difficulty, and congruency relative to the accuracy, reaction time, and
confidence outcomes for structural judgments. Blocks 2 and 3 established a basis for comparing
reaction times associated with plausibility judgments for sentences that appeared outside or
inside of arguments that were valid or invalid, easy or difficult, and congruent, incongruent
without a warning signal, or incongruent with a warning signal.

Results

Evaluating Structure in Valid Syllogisms

A repeated-measures ANOVA was performed, treating Gender as a between-subjects
variable and Difficulty and Congruency as within-subjects variables. The dependent variables of
interest were accuracy, reaction time for making the validity judgment, and participants’ self-reported confidence about the validity judgment. Main effects of the independent variables are reported in Tables 15, 16, and 17. Appendix I summarizes non-significant results related to Difficulty and Congruency.

Table 15

*Main Effects of Difficulty on Accuracy, Structural Judgment Reaction Time and Confidence for Valid Syllogisms*

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficult</th>
<th>p &lt;</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.63</td>
<td>.55</td>
<td>.05</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time for structural judgment (ms)</td>
<td>13088.08</td>
<td>19011.75</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.30</td>
<td>3.93</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 16

*Main Effect of Congruency on Accuracy for Valid Syllogisms (Study 3)*

<table>
<thead>
<tr>
<th>Congruency</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU</th>
<th>p &lt;</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.83</td>
<td>.30</td>
<td>.64</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>

*Mauchly’s test of sphericity showed that the variances were significantly different (p<.05). The Greenhouse-Geisser degrees of freedom and p-value were referred to.
Table 17

Main Effect of Gender on Reaction Time for Structural Judgment for Valid Syllogisms (Study 3)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Reaction Time for structural judgment</th>
<th>$p$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>19450.51</td>
<td>.01</td>
<td>ns</td>
</tr>
<tr>
<td>Female</td>
<td>12649.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accuracy. Difficulty, $F(1, 52) = 6.21$, $p<.05$, $\eta^2 = .01$, and Congruency, $F(2, 104) = 65.74$, $p<.001$, $\eta^2 = .22$ had main effects on accuracy and their interaction was also significant, $F(2, 104) = 5.74$, $p<.01$, $\eta^2 = .02$ (see Figure 23).

Figure 23

Interaction Effect of Difficulty x Congruency on Accuracy for Valid Syllogisms (Study 3)

Participants were more accurate on easy trials than difficult ones. Post hoc testing revealed that the accuracy associated with each congruency condition was unique; thus, the congruent
condition (BBB), with 83% accuracy, was substantially higher than the incongruent condition without (BBU) and with the warning signal (UBU) which were associated with 30% and 63% accuracy, respectively. The warning signal afforded a distinct boost to accuracy relative to the non-warning signal incongruent condition, regardless of the difficulty level of the syllogism. Post hoc testing of the Difficulty x Congruency interaction showed that belief-content conflict afforded the most substantial boost to accuracy in easy rather than difficult syllogisms, a finding that replicates those in Study 2.

Reaction time for structural judgment. Participants spent significantly longer solving difficult syllogisms compared to easy ones, $F(1, 52) = 24.42, p<.001, \eta^2 = .06$, thereby replicating the effect of Difficulty level on reaction time that was reported in Study 2. Male participants took significantly longer than female participants to solve syllogisms, regardless of their difficulty level, $F(1, 52) = 12.25, p<.01, \eta^2 = .04$. Although it appears that participants spent more time on incongruent syllogisms than congruent ones, this difference was not quite statistically significant ($p=.06$) like it was Study 2.

Confidence. Participants were significantly more confident about their judgments of easy syllogistic structure than difficult syllogistic structure, $F(1, 52) = 19.58, p<.001, \eta^2 = .05$.

Evaluating Structure in Invalid Syllogisms

The effects of Gender, Difficulty, and Congruency of invalid syllogisms on the accuracy, length of time to judge the structure of the arguments, and associated confidence of participants’ reasoning were tested by way of a repeated-measures ANOVA. Main effects of the independent variables are reported in Tables 18, 19, and 20. Appendix I summarizes non-significant results related to Difficulty and Congruency.
Table 18

Main Effects of Difficulty on Reaction Time and Confidence for Invalid Syllogisms

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Reaction Time for structural judgment (ms)</th>
<th>Confidence</th>
<th>$p$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>16454.27</td>
<td>4.19</td>
<td>.001</td>
<td>ns</td>
</tr>
<tr>
<td>Difficult</td>
<td>20583.21</td>
<td>3.74</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 19

Main Effects of Congruency on Accuracy and Reaction Time for Invalid Syllogisms

<table>
<thead>
<tr>
<th>Congruency</th>
<th>Accuracy</th>
<th>RT for structural judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBU</td>
<td>.83</td>
<td>16893.31</td>
</tr>
<tr>
<td>BBB</td>
<td>.30</td>
<td>17792.04</td>
</tr>
<tr>
<td>UBB</td>
<td>.53</td>
<td>20870.86</td>
</tr>
</tbody>
</table>

*p* Mauchly’s test of sphericity showed that the variances were significantly different ($p<.05$). The Greenhouse-Geisser degrees of freedom and p-value were referred to.

Table 20

Main Effect of Gender on Reaction Time for Invalid Syllogisms

<table>
<thead>
<tr>
<th>Gender</th>
<th>RT for structural judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>22948.94</td>
</tr>
<tr>
<td>Female</td>
<td>14088.53</td>
</tr>
</tbody>
</table>

$\text{p} < .001$ ns
Accuracy. The Congruency of the conclusion in invalid syllogisms was found to significantly impact participants’ accuracy for determining whether the conclusion followed necessarily from the premises of the argument, $F(2, 104) = 56.54, p<.001, \eta^2 = .20$ (see Figure 24). In particular, participants were most successful (83%) in the congruent condition (BBU) when the conclusion’s implausibility matched the syllogism’s invalid structure. They were least successful (32%) when the conclusion’s plausibility did not match the structure (BBB). Finally, the warning signal condition where implausible content preceded a plausible conclusion (UBB), was associated with an accuracy level (53%) that was significantly higher than the BBB condition, yet not on par with the BBU condition.

Figure 24

*Main Effect of Congruency on Accuracy for Invalid Syllogisms (Study 3)*
Reaction time for structural judgment. Participants’ Gender, $F(1, 52) = 18.05$, $p<.001$, $\eta^2 = .05$, as well as the Difficulty of the argument, $F(1, 52) = 16.49$, $p<.001$, $\eta^2 = .02$, and Congruency between the conclusion and structure of the argument, $F(2, 104) = 5.50$, $p<.01$, $\eta^2 = .02$ (see Figure 25), significantly impacted the amount of time spent reasoning about invalid syllogisms. Male participants spent significantly longer reasoning about the syllogisms than their female counterparts, and participants generally spent more time reasoning about difficult than easy syllogisms. Post hoc testing revealed that the warning signal condition (UBB) was associated with significantly longer reaction times than the two other Congruency conditions.

Figure 25

*Main Effect of Congruency on Reaction Time for Invalid Syllogisms (Study 3)*
Confidence. Participants were significantly more confident about their structural judgments of easy compared to difficult invalid syllogisms, $F(1, 52) = 38.61, p < .001, \eta^2 = .06$.

Evaluating Syllogistic Structure - Summary of Findings and Preliminary Interpretations

Valid syllogisms. A number of main effects and one key interaction from Study 2 were replicated by Study 3. As expected the main effect of Congruency on accuracy that was found in Study 2 was replicated in this study in that statistically distinct levels of accuracy were found to exist for the congruent (BBB), incongruent minus the warning signal (BBU), and warning signal (UBU) conditions. Participants were (a) most accurate for the BBB condition, (b) least accurate for the BBU condition, and (c) relative to the BBU condition, the addition of a warning signal (UBU) greatly improved accuracy. The main effects of Difficulty on accuracy, reaction time, and confidence from Study 2 were also replicated. Participants were more accurate, spent less time reasoning, and reported greater confidence about their judgments of easy compared to difficult syllogisms. The interaction between Congruency and Difficulty for accuracy that was reported in Study 2 was replicated in Study 3. The warning signal condition (UBU) was found to boost accuracy most for easy as compared to difficult syllogisms. On the cusp of being replicated was the main effect of Congruency on reaction time that was found in Study 2, that incongruent syllogisms (BBU and UBU) were reasoned over for substantially more time than congruent (BBB) ones. This finding was significant in Study 2 ($p < .05$) and nearly significant in Study 3 ($p < .06$).

One main effect and one interaction from Study 2 were not replicated in Study 3. In Study 2 participants were found to be the most confident about their judgments of congruent (BBB) syllogisms compared to incongruent ones (BBU, UBU), and the presence of a warning signal was found to boost confidence substantially for syllogisms containing an implausible
conclusion compared to lacking this cue. Study 3 did not replicate this finding. The interaction between Difficulty and Congruency for reaction time that was reported in Study 2 was not produced in Study 3; according to Study 2 more time was spent on incongruent trials (BBU and UBU) than congruent trials for easy syllogisms, and less time was spent on the warning signal condition (UBU) compared to the BBB and BBU conditions for difficult trials. There were no significant differences between difficult and easy syllogisms according to Congruency according to Study 3.

*Invalid syllogisms.* Similar to the findings in Study 2, it was found that distinct levels of accuracy existed for the three Congruency conditions. Specifically, the greatest accuracy occurred when there was congruency between the conclusion and the structure (BBU) and the lowest accuracy was observed when there was a mismatch between the two (BBB). The detriment of this mismatch on accuracy seemed to be remedied when a warning signal preceded a plausible conclusion (UBB). The finding from Study 2 that participants spent more time reasoning about the warning signal condition than the other two Congruency conditions was replicated in Study 3. In terms of Difficulty level, Study 3 found that participants reasoned for longer durations and with more confidence about easy syllogisms than difficult ones, thereby replicating two of the findings from Study 2. What failed to be replicated in Study 3 were the findings that easy syllogisms were resolved more successfully than difficult ones; Study 3 showed equal accuracy for the two difficulty levels. Also, participants were not found to be more confident about their judgments for the congruent condition (BBB) compared to the two incongruent ones, as was reported in Study 2; instead, equal confidence levels were reported by participants across the Congruency conditions.
Evaluating Semantic Plausibility

Total time for 3-part syllogism. The total amount of time that participants spent evaluating the plausibility of an argument was calculated by adding the reaction times associated with the major and minor premises as well as the conclusion of individual arguments. This value was then divided by the number of words in the syllogism thereby creating a set of weighted values that were included in the ensuing analysis. A repeated-measures ANOVA was completed to determine whether the amount of time participants spent on the semantic judgment task differed according to the Difficulty level of the syllogistic argument (i.e., in that it contained three affirmative assertions or one negative and two affirmative assertions), as well as the relative Plausibility Pattern in the syllogism (BBB, BBU, UBB, UBU). Validity of the argument was not considered as an independent variable because participants were instructed to attend to the content rather than the structure of the argument.

Two significant main effects (see Tables 21 and 22) and one interaction were found related to the amount of time participants spent judging the empirical validity of three-part syllogisms.

Table 21

Main Effect of Difficulty Level on Semantic Evaluation Reaction Time for Whole Argument

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficult</th>
<th>( p &lt; )</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Time for semantic judgment of whole argument (ms)</td>
<td>688.48</td>
<td>913.47</td>
<td>.001</td>
<td>ns</td>
</tr>
</tbody>
</table>
Table 22

*Main Effect of Plausibility Pattern on Semantic Evaluation Reaction Time for Whole Argument*

<table>
<thead>
<tr>
<th>Plausibility Pattern</th>
<th>BBB</th>
<th>BBU</th>
<th>UBB</th>
<th>UBU</th>
<th>p &lt;</th>
<th>Mauchly's test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Time for semantic judgment of whole argument (ms)</td>
<td>632.91</td>
<td>749.41</td>
<td>885.39</td>
<td>809.66</td>
<td>.001</td>
<td>*</td>
</tr>
</tbody>
</table>

* Mauchly’s test of sphericity showed that the variances were significantly different (p < .05). The Greenhouse-Geisser degrees of freedom and p-value were referred to.

In terms of the Difficulty level of the argument, difficult arguments containing a negative proposition (No A are B) were associated with longer periods of evaluation than easy arguments containing three affirmative statements, $F(1, 52) = 74.41$, $p < .001$, $\eta^2 = .06$. The Plausibility Pattern also impacted reaction time, $F(3, 156) = 10.02$, $p < .001$, $\eta^2 = .04$ (see Figure 26), and post hoc testing revealed that the semantic evaluation of arguments containing three plausible statements (BBB) took significantly less time than the semantic evaluation of arguments containing implausible statements (UBB and UBU, specifically).
Finally, there was an interaction between Difficulty x Congruency, $F(3, 156) = 18.94, p<.001, \eta^2 = .04$ (see Figure 27). Post hoc testing revealed that when evaluating the plausibility of easy arguments containing a particular affirmative major premise, participants took the longest to judge those of form UBB compared to BBB, BBU, and UBU. In other words, participants took the longest when the major premise was an implausible particular statement (e.g., Some babies are adults) than when it contained three plausible propositions, an implausible conclusion only, or an argument containing an implausible major premise and conclusion. In the case of difficult arguments containing a negative major premise, the most time consuming plausibility pattern was BBU; participants grappled longest with arguments starting with plausible and universally disqualifying sentences (e.g., No carrots are pink) and ending with an implausible conclusion (e.g., Some flowers are carrots).
Figure 27

*Interaction Effect of Difficulty x Plausibility Pattern on Reaction Time for Semantic Evaluation of Whole Argument*

![Graph showing the interaction effect of Difficulty x Plausibility Pattern on Reaction Time for Semantic Evaluation of Whole Argument.](image)

*Time to judge conclusion.* The amount of time that participants spent determining whether or not the conclusion statement in a syllogism was plausible in real-life was calculated on each trial. Weighted values were calculated across the conditions to account for slight variations in the number of words per conclusion. These weighted values were then included in the ANOVA model, testing for differences in reaction time according to the Difficulty level and the Pattern of Plausibility of the content in the argument. Validity of the argument was not considered as an independent variable because participants were instructed to focus in a piecemeal manner on the individual components of the argument and were not asked to holistically evaluate the structure of the argument. Difficulty and Plausibility Pattern yielded
main effects which are reported in Tables 23 and 24. The interaction between Difficulty and Plausibility Pattern was not significant.

Table 23

Main Effect of Difficulty on Semantic Evaluation Reaction Time for Conclusion Statement Only

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Easy</th>
<th>Difficult</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>406.45</td>
<td>470.68</td>
<td>.05</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Time for semantic judgment of conclusion only (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 24

Main Effect of Plausibility Pattern on Semantic Evaluation Reaction Time for Conclusion Statement Only

<table>
<thead>
<tr>
<th>Plausibility Pattern</th>
<th>BBB</th>
<th>BBU</th>
<th>UBB</th>
<th>UBU</th>
<th>$p &lt;$</th>
<th>Mauchly’s test of sphericity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>432.70</td>
<td>603.49</td>
<td>123.29</td>
<td>594.79</td>
<td>.001</td>
<td>*</td>
</tr>
<tr>
<td>Reaction Time for semantic judgment of conclusion only (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Mauchly’s test of sphericity showed that the variances were significantly different ($p<.05$). The Greenhouse-Geisser degrees of freedom and p-value were referred to.

Participants spent significantly longer evaluating the plausibility of conclusions that were embedded in difficult syllogisms compared to easy ones, $F(1, 107) = 5.36, p<.05, \eta^2 = .01$. In other words, conclusions that were preceded by a major premise that was a negative proposition (No A are B) were considered for longer periods of time than those that were preceded by a major premise that was an affirmative statement (Some A are B). In terms of the Plausibility
Pattern, participants took an equal amount of time to evaluate the plausibility of conclusions that were embedded in arguments of the form BBB, BBU, and UBU and the least amount of time to evaluate conclusions embedded in arguments of the form UBB, $F(3, 321) = 64.33, p<.001, \eta^2 = .22$ (see Figure 28).

Figure 28

_Main Effect of Plausibility Pattern on Reaction Time for Semantic Evaluation of Conclusion_

![Reaction Time for Semantic Evaluation of Conclusion](image)

**Plausibility Pattern**

_Difference between time to evaluate conclusion inside and outside of argument._

Participants spent less time evaluating the plausibility of conclusion statements when they were presented in the company of other propositions than when they reviewed these statements independently outside of the argument context. Counter-balancing was used in this study so it is unlikely that this effect can be explained based on the order of presentation or a varying level of
conclusion novelty between the isolation and contextualized conditions. Rather, it seems that the argument context itself is what may explain participants’ diminished attention to the conclusion.

By subtracting the amount of time that participants spent evaluating the proposition inside of the argument from the amount of time they spent evaluating that same proposition outside of the argument (RTout – RTin), a value was produced that represents the time that was forfeited by participants once the task divided their attention. A small value suggests that despite the fact that participants were reviewing a conclusion statement in the company of other, potentially distracting propositions, a small amount of time was forfeited relative to the time expended when that statement was reviewed independently of the other two premises. A large number on the other hand, shows that the conclusion statement was not able to hold participants’ attention effectively once it was placed in the mixed company of the other premises, thus there was a great trade-off in time.

If the argument contains a negative assertion (No A are B) as its major premise, does this mean there is a great attention trade-off and that the conclusion is no longer attracting attention compared to an easier argument context where the conclusion is preceded by affirmative statements? How does the Pattern in which Plausible/Implausible content is distributed within an argument impact the conclusion’s ability to hold the reasoners’ attention while they evaluate an argument’s plausibility? These two questions led to an ANOVA model in which Difficulty and Plausibility Pattern were treated as within-subjects variables. The argument’s Validity was not factored in because participants were not being asked to evaluate the structure of the argument. No main effect of Difficulty level was detected, nor was there an interaction between the two independent variables.
The Pattern of Plausibility in the argument was found to significantly affect the trade-off in reaction times for the semantic evaluation of conclusions inside and outside of arguments, $F(3, 321) = 10.33$, $p<.01$, $\eta^2 = .04$. The trade-off was fairly low, meaning that the conclusion statement held its own when included in an argument so long as the argument was of the form BBB, BBU, or UBU. The trade-off was substantial, meaning that participants were distracted by the components of the argument and focused much less on the conclusion when it was included in an argument of the form UBB (see Figure 29).

Figure 29

*Main Effect of Plausibility Pattern on Trade-off between Reaction Time for Semantic Evaluation of Conclusion Inside Versus Outside of an Argument*

In general, participants took more time to determine whether the content of arguments reflected reality when the arguments opened with an implausible statement (UBB and UBU) than when they were wholly believable (BBB). When it came to the piecemeal semantic evaluation of
arguments’ conclusions, implausible conclusions (BBU and UBU) as well as conclusions in wholly believable arguments (BBB) were evaluated for longer durations than plausible conclusions found in arguments that started out with an unrealistic major premise (UBB). Participants’ natural inclination was to spend less time focusing on conclusions when these propositions appeared in the context of two other statements, but, this inclination led participants to be least distracted away from the conclusion when the conclusion was implausible (BBU and UBU) or preceded by two wholly plausible statements and itself plausible (BBB). Participants’ attention was most likely to be diverted from the conclusion when the conclusion was believable and preceded by an unlikely major premise. These findings suggest that implausible content is a stumbling block in that it requires more effortful consideration. In the case of the UBB condition, the implausible opening line of the argument seemed to detract from a lengthy consideration of the realism of the conclusion.

Syllogisms containing affirmative statements (e.g., Some A are B) are easier than their counterparts containing a negative statement (e.g., No A are B). Not surprisingly, when participants were asked to determine whether a statement represented the real world, they took substantially longer to evaluate the entire argument as well as the conclusion statement when that statement was a disqualifying comment as compared to an affirmative assertion about the world. When the content of the argument was wholly affirmative in that it contained Some and All statements, the most time-consuming form of syllogism was one in which the conclusion was plausible yet preceded by an implausible major premise. In contrast, the most time-consuming conclusion in a difficult argument containing a disqualifying negative premise was one that was implausible, yet preceded by a plausible opening line.
In a nutshell, participants spent more time evaluating the meaning of whole arguments that deny rather than affirm and that are highly implausible rather than plausible. They also conducted the lengthiest evaluations of conclusions associated with denial rather than affirmation; the shortest semantic evaluations were for conclusions that were believable having been preceded by an implausible major premise. Interestingly it was in this condition (UBB) where the greatest trade-off was observed between the reaction time inside and outside of the argument for the semantic evaluation of the conclusion. That is to say that, scant attention was given to believable conclusions when they were preceded by more salient, implausible content in the argument. This shifting in attention whereby the conclusion was considered for less time in order to accommodate premise-oriented evaluation was not observed to the same degree in the BBB, BBU, and UBU conditions. Presumably this is because in the BBB condition the premises were plausible and therefore no more interesting/self-relevant than the conclusion. In the BBU condition the implausibility of the conclusion may have anchored participants’ attention to this part of the argument given the irrelevance of the plausible premises. Finally, in the UBU condition, participants faced two statements that may have competed for attention. They appeared to divide their attention between the conclusion and the major premise, a process that did not lead to conclusion neglect. Conclusion neglect was sponsored only in the UBB condition in which the neutrality of the conclusion afforded participants with the freedom to attend to something more pressing – the implausible major premise.
CHAPTER 10: DISCUSSION

Studies 1A, 2, and 3 replicated the belief bias effect (Janis & Frick, 1943; Evans, Barston & Pollard, 1983), an effect that stems from the relationship between conclusion content and syllogistic structure. Specifically, participants were highly accurate when judging the syllogistic structure of congruent arguments and highly inaccurate when judging incongruent arguments. Research by Evans, Newstead, and Allen (1994) as well as Goel and Vartanian (2011) established that the belief bias effect can be reduced, if not entirely eliminated, under the right conditions (e.g., verbal instructions that explain the logical meaning of “some” or the inclusion of negative emotional content within the argument). My dissertation extends what is known about belief bias by revealing that belief-content conflict within the premises of an argument is another factor that can diminish this effect, and that the efficacy of this cue depends on where it is located in the argument as well as the syllogism’s difficulty level. How long participants took to judge a given syllogism as well as the confidence they attributed to this judgment were observed in addition to accuracy; these variables shed light on the processes that may be prompted by belief-content conflict. Studies 1B and 3 delved further into the processes that may derive from belief-content conflict by investigating how reasoners judge the empirical validity of syllogistic content when this cue is present and absent.

Belief-Content Conflict Attenuates Belief Bias

Valid Trials

In valid trials that were incongruent, the addition of belief-content conflict in the introduction of the argument was observed to boost accuracy, an effect that was first documented in Study 1A and then replicated in Studies 2 and 3. The effect of belief-content conflict seemed to be optimized when this cue was presented as the minor rather than the major premise because
from this position it produced the most substantial boost in accuracy (1A), was most likely to have its implausible content noted by participants (1B), and seemed to diminish participants’ sensitivity towards the conclusion’s implausibility (1B). Belief-content conflict also seemed to be more effective at attenuating belief bias in easy rather than difficult arguments. Specifically, participants judged the logical structure of incongruent syllogisms most accurately when the arguments in which belief-content conflict was placed were easy (containing affirmative propositions) rather than difficult (containing negative propositions). This finding was documented in Study 2 and replicated in Study 3.

Clearly, critical features of the syllogism determine the efficacy of belief-content conflict as a cue that attenuates belief bias in valid syllogisms. How the cue functions is a question that we can attempt to answer by looking at the reaction time and confidence measures. When faced with an implausible conclusion, the hallmark of incongruence in valid arguments, participants deliberated at length before rendering a judgment. The addition of belief-content conflict did not elongate or diminish response time in Studies 1A or 3, although there was evidence in Study 2 that the addition of belief-content conflict into incongruent syllogisms substantially decreases response times for difficult syllogisms that contain negative propositions. It is not possible to draw clear-cut conclusions about how the addition of belief-content conflict to an incongruent argument impacts participants’ confidence about their ensuing syllogistic judgment. Although Study 2 showed that belief-content conflict trials boosted participants’ confidence about their syllogistic evaluations, we must err on the side of caution and assume that confidence is not improved or diminished by belief-content conflict given that two other studies led to null effects of the warning signal on confidence. Specifically, participants were not overly confident when judging incongruent trials, regardless of whether belief-content conflict was present or absent
(Study 1A) or whether this cue was embedded in easy affirmative or difficult negative propositions (Study 3).

Taken together the findings imply that System 2 is implicated in participants’ consideration of the belief-content conflict trials. Not only are accurate syllogistic judgments promoted when this cue is present, this effect is most pronounced when syllogistic structure is easy. In this case effortful System 2 processing is the most likely to pay off through the derivation of a conclusive (and accurate) answer. Although reaction time was not greater for incongruent trials with a warning signal than it was for those without one, the amount of time that participants spent on these incongruent trials was high relative to congruent trials. This suggests that belief-content conflict may have changed the specific System 2 analysis engaged during reasoning, rather than the engagement of System 2 altogether. Recalling that belief-content conflict was most effective when embedded as the minor premise (1A) and that it appeared to diminish participants’ focus on conclusion content (1B) without compromising accuracy or diminishing reaction time implies that part of its role as a cue is the redistribution of System 2 resources. Elaborating on this idea, if the cue’s inclusion is associated with diminished emphasis on conclusion content, then it stands to reason that individuals expend cognitive effort by assessing qualities other than argument content or by applying alternate analytic tools than when this cue is absent. Thus, the cue may prompt critical analysis that is different in kind but not in degree. The types of analytic strategies that this cue might prompt were discussed at length in Chapter 6 (recall: forwards rather than backwards reasoning; complex and multidimensional versus binary relational thinking; and hypothetical thinking).

In line with the finding that belief-content conflict seems to inspire a shifting of gears within System 2, future studies should explore the types of reasoning strategies that may be
inspired by the presence of such a cue in syllogistic reasoning problems. For instance, researchers might take inspiration from Feather’s (1964) observation that belief bias predominates in reasoners with low tolerance for inconsistency. In line with this finding, one might aim to determine whether belief-logic conflict attenuates belief bias by inspiring a more tolerant frame of mind in reasoners. If reasoners are prompted by belief-content conflict to engage hypothetically with syllogistic content they might become less vulnerable to belief-logic conflict. Additionally, given that Piper (1985) showed that syllogistic reasoning performance could be boosted by constructing a problem-solving context that would capitalize on reasoners’ familiarity with fictional narrative worlds, it would be interesting to pursue in future studies the question of whether the belief-content conflict cue inspires reasoners to engage more liberally with the problem by imagining a surreal context that is more accepting of implausible content.

Although Study 1B did not support the interpretation that belief-content conflict actually prompts reasoning about syllogistic structure, this possibility cannot be ruled out because the task set in Study 1B, gauging participants’ post-reasoning interest in syllogistic structure, may have been limited in its ability to capture their in vivo consideration of syllogistic structure. As such, a third direction that is suggested by this research is one that ought to examine how syllogistic structure is considered during reasoning when belief-content conflict is present and absent. An eye-tracking paradigm in conjunction with a think-aloud approach would be particularly useful for examining the ways in which belief-content conflict may shape reasoning. This dual approach could test two relevant hypotheses. Firstly, that premise-based belief-content conflict promotes balanced structural analysis whereby equal attention is meted out across the argument’s three propositions (rather than the conclusion being prioritized). Secondly, eye-tracking techniques in combination with a think-aloud paradigm could reveal whether belief-
content conflict encourages forwards reasoning whereby the premises are used to generate a conclusion.

Invalid Trials

As was observed for valid trials, adding belief-content conflict to the premises of invalid arguments attenuated belief bias and the specific position of this cue impacted its efficacy. The key differences between valid and invalid trials included the fact that the location of the cue determined whether or not the effect would happen (invalid trials) rather than how pronounced this effect would be (valid trials). Secondly, in valid trials the difficulty level of the syllogism was a limiting factor that prevented belief-content conflict from ameliorating accuracy in challenging arguments, yet this factor did not impact the efficacy of the belief-content conflict cue in invalid trials. In Study 1A the belief-content conflict cue was only effective when presented as a major premise; in this case it was observed to lower accuracy when added to congruent arguments. Study 1B showed that when presented as a major premise in invalid congruent trials, the warning signal made participants less sensitive to the implausibility of the conclusion. When belief-content conflict was embedded as a minor premise in invalid congruent trials it was not effective in terms of diminishing accuracy (1A) nor did it diminish participants’ sensitivity to the conclusion’s implausibility (1B).

The question of why the minor premise was not an optimal location for embedding belief-content conflict warrants some consideration. Study 1B ruled out possible explanations for why this cue was ineffective in this position. The efficacy of belief-content conflict was determined by comparing two conditions, BUU and BBU; if BUU trials were judged with significantly less accuracy than BBU trials, it would have been concluded that the cue was attenuating belief bias successfully. There was no significant decrease in accuracy however. This
might have been because the BBU trial did not induce belief bias, an explanation that was ruled out in Study 1B upon observing that participants were sensitive to the implausibility of the conclusion, and therefore likely to be biased by it. An alternate explanation, that the cue failed because it was not noticed by participants, was similarly ruled out in Study 1B because participants were observed to be sensitive to the implausible minor premise.

An account that seems the most likely is one that assumes that participants are implicitly aware of syllogistic structure (Morsanyi & Thompson, 2012) which would make them sensitive to the congruency between a problem’s structure and content. Compared to an incongruent trial, a congruent trial represents less of a conflict and presumably exerts less pressure on participants to engage in effortful and non-heuristic elaboration of the problem since their intuitive ‘feeling’ about the problem’s structure matches their surface evaluation of its content. In this instance of congruency, it would therefore make sense that a cue that warns participants not to rely on belief-driven heuristics ought to be presented as early as possible in the problem to allow for maximum exposure. The chance that a conflict cue could disrupt participants’ natural tendency to deny validity to an invalid and implausible conclusion ought to be greatest for cues presented for long durations rather than short ones. As such, when belief-content conflict appears in the minor premise, it may be a signal that challenges participants’ use of a belief-driven heuristic strategy, but one that is presented too briefly before participants view the conclusion and render their ultimate judgment. If however the conflict cue is presented early in the problem as the major premise, the exposure time may be great enough to allow the conflict cue to influence participants, thereby disrupting a belief-driven style of judgment that was encouraged by participants’ intuitive sensitivity to congruency. Reaction time analyses comparing congruent invalid trials with and without the belief-content conflict cue revealed that processing was fast in
either case, a finding that suggests that the belief-content conflict cue leads to a shifting of gears within System 1 rather than an initiation of more effortful, System 2 based reasoning. Accordingly, participants may apply structurally-driven heuristics as opposed to content-driven ones in congruent trials in which the warning signal appears as the major premise.

The belief-content conflict cue was most effective in invalid syllogisms when presented as the major premise, a finding that seems at first glance to be inconsistent with the effect it has on invalid trials. However, it is important to note that the cue played fundamentally different roles in the valid and invalid trials – in the former it disrupted nay-saying in the context of incongruency whereas in the latter it disrupted nay-saying in the context of congruency. If participants were implicitly aware of the congruency between content and syllogistic structure, it would presumably take greater impetus (read: earlier exposure) to overturn nay-saying than if they were implicitly aware that there was a mismatch between syllogistic content and structure. In the former case, early exposure to belief-content conflict in the major premise might be the only way to overturn belief bias (by reducing conclusion focus as per Study 1B). In the latter case, it might not be necessary to immediately deploy the belief-content cue to permit maximum exposure and rather, it might be more successful when nested next to the conclusion in a position from which to diminish participants’ sensitivity to the conclusion. Study 1B emphasized the potential role of belief-content conflict in diminishing sensitivity to conclusion content – specifically, implausible propositions that were rated as equally unlikely during pre-testing were less likely to be identified by participants implausible conclusions when they were immediately preceded by an implausible minor premise (BUU) as opposed to an implausible major premise and a plausible minor premise that was plausible (UBU).
When belief-content conflict was inserted into the major premise of incongruent invalid syllogisms, accuracy was substantially greater than it was in incongruent trials without this cue (Studies 2 and 3) and attention was actively drawn away from the conclusion by the major premise, as evidenced in Study 3. Thus, belief-content conflict has the same effect on accuracy in invalid and valid trials when the condition of incongruency exists. In terms of underlying processes that might be inspired by this belief-content conflict in invalid syllogisms, the reaction time and confidence measures reveal some interesting patterns. In the case of congruent arguments, participants spent less time when the cue was present than absent; in either case they reported low confidence. The short reaction time implicates System 1 heuristic processing; however, the low level of accuracy rules out the likelihood that this type of processing was belief-oriented and suggests that some sort of systematic effort was made to evaluate the argument. If a heuristic strategy driven by syllogistic content was applied then accuracy should have been inflated rather than diminished because the implausible content ought to have encouraged a judgment of ‘invalid’. Perhaps participants applied a rule-based method of logical analysis to the syllogism and this was a System 1 driven process. This interpretation would accord with Handley et al.’s (2011) assertion that Systems 1 and 2 are equipped to shallowly evaluate semantic content and logical structure. It is possible that belief-content conflict, therefore, encouraged the application of ‘fast’ logical processing. This would have ensured that accuracy decreased because syllogistic content was no longer being considered, nor were the parts of the argument being evaluated in a ‘slow’ methodological manner that might have sustained accuracy once participants’ reliance on conclusion content was defeated.

In terms of incongruent invalid trials, longer reaction times were observed when belief-content conflict was present than absent (Studies 2 and 3). Although this finding was replicated,
the same cannot be said for confidence; participants reported lower confidence in Study 2 when the cue was present than absent, yet there was no apparent difference in confidence in Study 3 regardless of whether the cue was present or absent in easy or difficult trials. Recalling that there were no differences in accuracy between easy and difficult trials, the findings imply that System 2 was activated rather than reoriented by belief-content conflict in invalid, incongruent trials. It seems reasonable to conclude this because in the strictly incongruent trial (BBB), accuracy and reaction time were both low and the shift that was inspired by the presence of belief-content conflict cue led fast and imperfect System 1 processing to be replaced by slow and methodical System 2 activity.

The hypothesis that belief-content conflict inspires a shifting of gears within System 1 for congruent arguments implies that future research should follow two directions. For one thing, measures that are sensitive to the application of System 1 structural heuristics should be incorporated in order to ascertain whether belief-content conflict in fact prompts reasoners to apply shortcuts that operate on structural features of the problem as opposed to content-based ones. In order to determine whether structural heuristics are at work, implicit measures such as those applied in Morsanyi and Handley’s (2012) research ought to be employed. Recalling that Morsanyi and Handley assessed participants’ intuitive evaluations of syllogisms by asking them to rate their ‘liking’ of conclusions, an approach that is consistent with Winkielman and Cacioppo’s (2001) processing fluency hypothesis, it is clear that the methodologies applied throughout this dissertation are more appropriate to the detection of System 2 processes. The procedure of asking participants to explicitly judge argument validity is inherently more sensitive to System 2 reasoning processes and less sensitive to any structural assessment that may be going on implicitly within System 1.
Similarly, asking participants to indicate whether they would prefer to review a Venn diagram as opposed to evidence substantiating argument content is a task that would have limited ability to detect whether participants have applied some form of System 1 structural heuristic. A more appropriate approach for testing the hypothesis that structural heuristics replace content-driven ones when belief-content conflict is presented in a syllogism could be to present a set of syllogisms that would result in obvious, structurally driven response biases (e.g., Chater & Oaksford’s 1999 min- or max- heuristic) if this argument feature were being attended to.

A final direction that is implied by the findings in this dissertation is one that highlights the importance of observing participants’ responses to belief-content conflict in the context of congruent and valid arguments. To this point it has only been ascertained that belief-content conflict attenuates belief bias in invalid, congruent arguments by decreasing artificially ameliorated performance on such trials. If a similar decrease in accuracy were observed when participants are asked to solve valid syllogisms embedded with both a believable conclusion and belief-content conflict in the premises, then the conclusion that such a cue prompts shifting within System 1 would be supported.

*System 1 versus System 2*

*Strategy shifting and activation.* Clearly, valid and invalid arguments represent distinct arenas for the belief-content conflict cue to play its role in syllogistic reasoning. In valid trials this cue is best deployed in easy trials, however in either trial type (easy or difficult) it seems to prompt a *shift in activity within System 2*. This shift may entail substitution of a default mode of reasoning, for instance the tendency to use the conclusion to generate premises, with a different analytic approach, for example the implementation of a forwards style of reasoning where the premises are used to derive a conclusion. The System 2 shift initiated by belief-content conflict
may represent more than the mere initiation of a thinking style that is commensurate with the
task at hand. Rather, the cue may optimize the reasoner’s use of logical tools so that in the end
the analysis they perform is the best possible test of logic despite the fact that multiple analytic
techniques could fit the bill and would qualify as being task-relevant. For instance, syllogistic
reasoning necessitates relational thinking and it could be that the cue inspires complex
multidimensional thinking whereby the three parts of the syllogism are considered in a balanced
manner as opposed to a binary style of thinking where the parts of the argument are juggled with
less sophistication.

In contrast, in invalid trials which are inherently more difficult than their valid
counterparts, difficulty level does not determine the magnitude of the effect of belief-content
conflict. Instead, in congruent invalid trials belief-content conflict seems to inspire a shift in
activity within System 1 whereby belief-orientation is replaced by a rudimentary consideration of
logical properties of the problem. Participants were observed to quickly process this trial type;
they also reported moderate levels of confidence which substantiates the proposed involvement
of System 1.

Finally, in incongruent invalid trials the presence of belief-content conflict seemingly
activates System 2. The difficulty of invalid trials likely encourages a low-cost, low-effort
heuristic style of evaluation. When these trials are congruent, a quality that participants may by
implicitly aware of (Morsanyi & Thompson, 2012), they are not motivated to conduct a thorough
examination of the argument unless a salient (read: implausible) statement is presented early in
the argument. In this instance participants shift their attention towards the logical structure of the
argument; however this is a half-hearted gesture that is heuristic rather than wholly analytic and
effortful. It makes sense that participants do not conduct an effortful logical assessment of the
argument because the congruency of these trials likely confers to participants the sense that a belief-driven solution ‘could be good enough’. In the case of incongruent invalid syllogisms, the belief-content conflict cue may be able to activate deep logical assessment (System 2) rather than shallow logical scanning (System 1) because participants implicitly appreciate that content and structure are discordant. Participants would therefore sense that a heuristic strategy may not pay off and therefore the effort and time required to logically evaluate the problem would be warranted and System 2 activity would result.

**Inhibition.** The evidence presented here supports the conclusion that some conditions essentially ‘press the gas pedal’ for the analytic system. This assertion might be tested by looking to see if there is activity in the rLPFC, a brain region that acts as a ‘sponsor’ for the analytic business of recognizing conflicts and orienting towards logical goals (Goel & Dolan, 2003). The question of whether these same conditions inspire a ‘pressing of the brake pedal’ for the heuristic system remains unanswered. Belief-content conflict may lead to System 1 inhibition, perhaps by way of rLPFC activity (Tsujii & Watanabe, 2009), but this would have to be tested in future studies. In the current studies belief-content conflict attenuated belief bias successfully, an effect that seemed to relate to content-blocking (i.e., participants’ sensitivity to implausible conclusion content was diminished by the belief-content conflict cue in Study 1B). Based on the conventional view that System 1 is a superficial processor of problem content whereas System 2 drives critical reflection and spatial manipulation of structural features of problems (e.g., Goel & Dolan, 2003), one might conclude that content-blocking is indicative of System 1 inhibition. De Neys and Franssens (2009) provided evidence that inhibition entails intuitive belief blocking. Specifically, participants who blocked content during syllogistic tasks were impeded on their subsequent recognition of this content in time-limited judgment tasks, implying that content
blocking necessitated an inhibition of System 1. According to Handley, Newstead and Trippas’s (2011) theory however, the findings reported by De Neys and Franssens may be insufficient evidence for the conclusion that System 1 inhibition occurs when content is blocked. Rather, their theory contends that both systems process content and structure, albeit at different rates and with more or less diligence. Thus, it is not possible to conclude that content-blocking necessarily implies System 1 blocking.

Regardless of whether content is considered intuitively by System 1 or analyzed in greater depth by System 2, what has been established by my research is that belief-content conflict can induce content-blocking for syllogistic conclusions. Blanchette and colleagues completed three studies that in combination may shed light on why belief-content conflict in the premises of syllogisms leads to content inhibition in the conclusion. Incidental (task-irrelevant) emotions and implicit arousal about syllogistic content compromise syllogistic reasoning whereas integral (task-relevant) emotions and explicitly experienced emotional responses to syllogistic content do not (Blanchette & Richards, 2010; Blanchette & Leese, 2011). Also of import is the finding derived from analyses of reasoning by eyewitnesses of terrorism who were shown to accurately solve terrorism-related problems so long as their mode of coping allowed them to explicitly inhibit their terrorism-related beliefs (Blanchette et al., 2007). Taken together these three studies indicate that there is a relationship between beliefs, implicit versus explicit experience, inhibition, and reasoning success. This set of findings suggests that beliefs that are explicitly experienced are more easily inhibited and less likely to compromise logical reasoning. Thus, applied to my research it seems as though the belief-content cue may push reasoners to consider all problem content in an explicit manner, thereby allowing this material to be readily inhibited when necessary. Inhibitory control is a key factor that has been shown to determine
whether belief bias occurs (Moutier et al., 2006), thus belief-content conflict may help reasoners to exert inhibitory control by rendering explicit the content that must be blocked.

Returning to the question of whether this content-blocking necessarily entails System 1 inhibition, I would argue that belief-content conflict as instantiated in the present experiments does not lead to System 1 inhibition. Belief-content conflict appears to induce an explicit (read: conscious) and task-relevant (read: goal oriented) treatment of content thus it seems that instead of System 1 being curbed by this cue, that it is more likely that System 2 is the locus of inhibition. Belief-content conflict may lead to a curbing of content elaboration by System 2 in an effort to make way for more situation-appropriate, goal-oriented analyses. This interpretation accords with Stollstorff, Vartanian, and Goel’s (2012) finding that belief-content conflict in the premises of spatial reasoning problems is associated with activity in the rLPFC, an area that acts as an analytic ‘sponsor’ by identifying conflicts and encouraging goal orientation by discouraging goal-discrepant content elaboration.

An fMRI paradigm would seemingly shed a lot of light on the presumed mechanisms underlying the relationship between belief-content conflict, the belief bias effect, and Systems 1 and 2. For instance, in the case of difficult incongruent arguments that contain an immediate warning signal, the instigation of System 2 processing that is implied by increased reaction times and decreased confidence, could be verified by looking for increased rLPFC and bilateral parietal activity which would represent, respectively, that conflict has been recognized and logical goals have been prioritized, and secondly, that spatial information is being manipulated. If System 1 processing were occurring then left temporal and ventral medial prefrontal cortex activity would be observed due to the predominance of semantic information seeking, and affective/limbic system involvement, respectively.
Although an fMRI paradigm would be enlightening, it is important to note that the data I have reported and the method of including belief-content conflict within the premises of syllogistic arguments necessitate a redefinition of what ‘counts’ as reasoning. A key weakness related to fMRI studies that was noted by Goel (2007) was that these studies tend to suppose that reasoning is about the conclusion. Clearly my data suggest otherwise; specifically they indicate that belief-content conflict within the premises can change the shape of System 2 reasoning, shift the gears within System 1, and even instigate System 2 processing where once heuristics predominated. As Goel (2007) pointed out, the fMRI paradigm looks at brain activity associated with reasoning which “is taken to occur with the presentation of the conclusion and its integration into the premises” (p. 436). This methodology needs to be adjusted because as it stands it is insensitive to reasoning if this process unfolds in a feet-first and premise-driven manner.

The Six Models of Reasoning

Six competing models of reasoning were described in the introduction. None of these models can account for the current set of findings but some speculation about how belief-content conflict in the premises may interface with these models is warranted nonetheless.

Selective scrutiny. Belief-content conflict in the premises was shown to impact syllogistic reasoning, a finding that challenges the selective scrutiny model (Evans, Barston & Pollard, 1983) which assumes a backwards direction of reasoning wherein premises are not considered prior to the conclusion. The fact that an implausible major premise was found to improve the accuracy of participants’ judgments of invalid trials with plausible conclusions (relative to wholly believable 3-part arguments, as per Study 3) directly challenges Evans et al.’s contention that logical deduction is driven by implausible conclusions. Clearly, a conclusion can be
plausible and participants can nonetheless engage in logical deduction because they have been
driven by disbelief about the premises.

*Misinterpreted necessity.* A key part of the misinterpreted necessity model (Evans et al.,
1983) is the assumption that participants will engage in a forward style of reasoning (as they
appeared to do in our studies) and that they will demonstrate belief bias on invalid trials but not
on valid ones. Belief bias was demonstrated in valid and invalid trials in our studies thus this
second assumption was violated. However, this effect was found to be more common in difficult
than easy trials which accords with Evans and Pollard’s (1990) and Newstead, Pollard, Evans,
and Allen’s (1992) update to the original model, one that explored the possibility that reasoners
can feel uncertain about more than indeterminate invalid syllogisms (i.e., difficult valid
syllogisms) and therefore fall prey to heuristic-driven processing. The misinterpreted necessity
model therefore receives some support, but the fact that this model does not consider how the
premises may contribute to participants’ perceived uncertainty is a weakness that would need to
be addressed in future inceptions of this account of reasoning. What is more, the uncertainty that
is being referred to in the misinterpreted necessity model relates to syllogistic structure and does
not pertain to the perceived empirical validity of problem content. Thus, the interaction between
certainty about structure and content requires further investigation by proponents of this model
given that the current findings imply that such an interaction may exist.

*Mental models and verbal reasoning theory.* These models are not directly challenged by
the findings I have reported, but they cannot fully account for the role that appears to be played
by belief-content conflict in the premises. Regardless of whether problem content is represented
in the form of visual tokens (Oakhill, Johnson-Laird & Garnham, 1989) or verbal units (Polk &
Newell, 1995), it seems likely that reasoners in my studies attempted to integrate the parts of syllogisms, generate tentative conclusions, and construct alternate models.

In terms of the mental models theory, the belief-content conflict cue may have provided the impetus for participants to endure with an analytic assessment of the syllogism despite the congruency of problem content and structure that may have made heuristic processing a tempting option (e.g., invalid syllogisms in Study 1A). However, as mentioned previously, Study 1B failed to produce evidence that belief-content conflict actually prompts an orientation towards syllogistic structure. Thus, the present experiments produced no evidence to the effect that belief-content conflict promotes mental modeling or token manipulation. Although the mental models approach presumes that reasoners proceed in a ‘feet first’ style of reasoning whereby premises are modeled before the conclusion, this account is limited in terms of how it may be applied to the current findings because it does not refer to the role played by the plausibility of the premises. The validation stage is said to be the key point at which logical deduction occurs; thus, re-appraisal and deductive work occurs when a tentative conclusion is deemed implausible, yet this model does not anticipate what may occur if a conclusion is plausible, yet preceded by an implausible premise.

Oakhill et al. (1989) identified difficulty level as a key factor in the mental models account of reasoning for this is what determines the outcome of a syllogistic reasoning trial. Specifically, the desire to engage in elaborated reasoning may be present, but if a problem is structurally complex because it entails multi-model construction, reasoning accuracy may be poor and presumably heuristic strategies may have been used because analytic reasoning was too demanding. This explanation of difficulty level was a key to the rationale for Studies 2 and 3 in which difficulty entered the design as an independent variable. Difficulty level was indeed found
to influence the outcome of reasoning trials as Oakhill et al. predicted; in terms of the belief-content conflict cue, syllogisms that were structurally complex because they contained negative assertions were the ones in which the cue was found to be only moderately successful at boosting accuracy. This effect of difficulty level was only found for valid syllogisms which may derive from the inherent difficulty of invalid trials to begin with. In invalid trials the belief-content cue was found to promote reasoning but only when it was presented in the major premise. How this cue could function despite the inherent difficulty of invalid trials is a question that remains unanswered if the mental models account is the only frame of reference from which to evaluate the current studies.

Polk and Newell’s (1995) verbal reasoning theory accounts for deduction by positing that linguistic skills are applied to reasoning problems and that syntax rules are used to manipulate verbal units and generate inferences. Unlike the mental models account, reasoning is not thought to be restricted to the last stage of problem-solving (validation), a perspective that is perhaps more accommodating to the current studies which demonstrated that the premises seem to impact reasoning from the earlier stages of comprehension and description. These stages are treated by the mental models account as relative ‘passive’ stages where ‘true reasoning’ is absent, yet it seems that the premises must be integrated in a logical fashion from the get-go if they are going to impact the ensuing syllogistic judgment rendered by participants. Thompson et al. (2003) proposed the modified verbal reasoning theory in which the concept of a response deadline was introduced. Although no measures were included in the current set of studies to assess participants’ perceived response deadlines between conditions, future studies might seek to determine whether the belief-content conflict cue leads to attenuated belief bias because it affords greater patience to participants. Thompson et al. noted that the response deadline is
different depending on the plausibility of the conclusion – reasoners are patient when the conclusion is believable and yet they find implausible conclusions less palatable and are therefore more rushed in their assessments; the question of how the plausibility of the premises interacts with that of the conclusion remains unanswered and warrants examination in future studies.

**MPT model.** Klauer, Musch and Naumer’s (200) account highlighted the issue of response bias and culminated in the multinomial processing tree model as an explanation for belief bias given this factor. They identified a number of types of response bias, including the threshold of evidence participants require before they will render a judgment, a threshold that was described as differing when participants look to confirm or disconfirm an argument. This discussion seems relevant to the current set of studies although future research would need to test whether or not the belief-content conflict perhaps manipulates reasoning by prompting one or the other type of testing. Possibly belief-content conflict acts on reasoners by encouraging analyses aimed at confirmation over disconfirmation, or vice versa. According to Klauer et al. (2000) it is the conclusion’s plausibility that determines which strategy is applied to a specific syllogism; given the current findings that highlight the importance of considering the premises’ plausibility as well as the limits put on reasoning due to difficulty of the argument, an update to the MPT model ought to discuss the interaction between premise and conclusion plausibility as well as the outcome of either strategy in easy versus difficult arguments.

Klauer et al. discussed the role of ‘certainty’ and ‘uncertainty’ at length. Over the course of 8 different experiments they demonstrated that participants’ perceived uncertainty about structure can be manipulated from the beginning of the syllogistic task (by giving them an estimation of the ‘odds’ that a given argument might be valid, e.g., 3 out of 4 arguments in this
group are invalid). By manipulating uncertainty about structure Klauer et al. showed that participants’ use of heuristic versus analytic strategies could also be manipulated. Uncertainty was a feeling that may have also been induced in participants who encountered belief-content conflict within my studies. A question that remains unanswered is how the uncertainty associated with implausible content might compare to the uncertainty associated with structural odds; clearly both forms of uncertainty are active on the reasoner from the outset of the argument, thereby acting on the process in a ‘feet first’ manner. Yet these types of uncertainty derive from different sources (content versus structure) and have different outcomes – uncertainty about structure leads to guessing whereas uncertainty about premise-based content in many conditions seems to attenuate belief bias and improve reasoning by discouraging this type of haphazard approach. Of course, less effortful styles of analysis (e.g., shallow yet logically-oriented scanning of congruent invalid trials with belief-content conflict) more akin to guessing were also demonstrated in some conditions within my studies. This type of haphazard, low-effort processing may have related more to participants’ intuitive awareness of low-stakes congruency than uncertainty about content and structure however.

Receiver operating characteristics. Dube, Rotello and Heit’s (2011) recent and hotly contested account of belief bias supposes that one’s general tendency to confirm an argument weighs differently on reasoners depending on whether they are assessing valid or invalid trials and therefore that conventional statistical procedures used by proponents of the competing reasoning models have led to methodological inadequacies being mistaken for belief bias. Focusing as they were on confirmation bias, aye-saying, Dube et al. argued that the ‘hit’ rate on valid trials does not share a linear relationship with the ‘false positive’ rate on invalid trials and that any effect labelled as belief bias is merely a Type 1 error. They applied a signal detection
approach to syllogistic reasoning by identifying argument strength as the operative variable determining whether liberal (aye-saying) or conservative (nay-saying) responding was more likely. Dube et al. went on to present evidence that confidence ratings and perceived argument strength were associated with different tendencies to affirm a conclusion. Klauer and Kellen (2011) countered with the observation that binary judgments of valid versus invalid structure are linearly distributed when plotted against each other, but that the use of confidence ratings by Dube et al. to discriminate between high- and low- confidence valid and invalid trials may have been what created the seemingly non-linear relationship. My studies were not designed to test Dube et al.’s account and as such it is not possible for me to weigh in conclusively on this debate. What little insight I can offer is drawn from my observation that belief bias (if it exists) or confirmation bias (if belief bias is a conflated effect) is being explained by models that refer to the status of the conclusion; yet I have demonstrated that bias (belief- or response- driven) can be attenuated when belief-content conflict is present at the right time within the premises. The mere effect that the premises appear to have on reasoning is on its own important to explain and something that conventional models cover in an inadequate manner when explaining syllogistic reasoning in a broad sense.

Extension beyond the Six Models

None of the models is able to fully account for the data I have reported. As such, the implication that I would draw from this conclusion is that a model that integrates several of these approaches is necessitated in order to account for why belief-content conflict attenuates belief bias by improving accuracy in incongruent arguments and diminishing it in the case of congruency. Furthermore, the observation that belief-content conflict attenuates belief bias, but that its ameliorative effect has limits, leads to a second implication – that reasoning context, the
key driver of the dual-processing system, is not bound. Certain contexts seemingly necessitate the activity of System 1 because they are too complicated to permit systematic, all-encompassing elaboration by way of System 2. Other contexts inspire enough immediate uncertainty about System 1 operations to warrant either a reorientation of System 1’s emphasis, or the altogether bridling of fast intuition in favor of effortful reasoning.

The six models each assume that a dual-processing system exists, yet the boundary between the two, which seems to be determined by problem context, needs to be examined in future studies. Seemingly, at the level of the individual there is an underlying logical capacity that resides in System 2. What my research suggests is that one’s motivation to sample from the logical toolbox as well as the logical tools that may be selected for the task at hand are driven by the self-relevance of argument content and the degree of conflict between one’s beliefs and an argument’s features. Whether the tapping of one’s logical capacities ends in an accurate syllogistic judgment depends not on the will of the reasoner so much as task demands. This assertion could be tested in future studies by incorporating other than syllogistic difficulty as an independent variable distinguishing between conditions that would or would not challenge the limited working memory resources that characterize System 2 processing. What is secondarily implied by my research is that awareness of logical structure may not be exclusively the product of System 2 processing. In fact, System 1 may allow for an intuitive awareness of belief-logic conflict, a role that is supported by research by Dijskterhuis (2006) and Morsanyi and Thompson (2012), and yet that has not been alluded to by current syllogistic reasoning models.
Conclusion

An integrated approach to the study of syllogistic content and structure led to the investigation of premise-based belief-content conflict and its impact on belief bias. The belief-content conflict cue effectively attenuated belief bias in incongruent valid and invalid trials, as well as congruent invalid trials. Its efficacy was found to depend on the difficulty level of the syllogism in which it was embedded as well as the location of its placement. Reaction time analyses were used to guide interpretations about the relative engagement of Systems 1 and 2. The key findings suggested that belief-content conflict activated System 2 for invalid incongruent trials which would otherwise have been processed using low-cost heuristic means due to their inherently difficult nature. On the other hand, it appeared that in valid trials the cue led to a redirection of System 2 resources such that specialized analytic strategies were applied in incongruent trials preceded by belief-content conflict compared to those lacking this cue. Finally, belief bias was successfully offset by belief-content conflict even in cases of congruency. In congruent invalid trials without this cue participants’ intuitive awareness of the content-structure match appeared to lead to low-cost, belief-based guesses; yet when presented as the major premise this conflict cue appeared to shift System 1 processing away from content and towards logical structure. Albeit less diligent than System 2 analysis, the shallow consideration of structural features may have been viewed as a safer bet than any shortcut aiming to capitalize on syllogistic content.
References


Appendix A

*Study 1A: Warm-up Syllogisms*

<table>
<thead>
<tr>
<th>Syllogism</th>
<th>Premises</th>
<th>Validity Type</th>
<th>% of Students who got it right during pre-testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Barbara</strong>&lt;br&gt;AAA</td>
<td>All A are B&lt;br&gt;All C are A&lt;br&gt;All C are B</td>
<td>Valid&lt;br&gt;Non-semantic</td>
<td>97%</td>
</tr>
<tr>
<td><strong>Barbara</strong>&lt;br&gt;AAA</td>
<td>All birds have feathers&lt;br&gt;All sparrows are birds&lt;br&gt;All sparrows have feathers</td>
<td>Valid&lt;br&gt;Semantic</td>
<td>98%</td>
</tr>
<tr>
<td><strong>Darapti</strong></td>
<td>All A are B&lt;br&gt;All C are A&lt;br&gt;Some C are B</td>
<td>Valid&lt;br&gt;Non-semantic</td>
<td>60%</td>
</tr>
<tr>
<td><strong>EO4</strong></td>
<td>No A are B&lt;br&gt;Some B are not C&lt;br&gt;Some C are A</td>
<td>Invalid&lt;br&gt;Non-semantic</td>
<td>63%</td>
</tr>
<tr>
<td><strong>EO4</strong></td>
<td>No frogs are cows&lt;br&gt;Some cows are not brown&lt;br&gt;Some brown animals are frogs</td>
<td>Invalid&lt;br&gt;Semantic</td>
<td>58%</td>
</tr>
<tr>
<td><strong>IOO1</strong></td>
<td>Some A are B&lt;br&gt;Some C are not A&lt;br&gt;Some C are not B</td>
<td>Invalid&lt;br&gt;Semantic</td>
<td>11%</td>
</tr>
</tbody>
</table>
### Study IA: Non-Semantic Test Syllogisms

<table>
<thead>
<tr>
<th>Syllogism</th>
<th>Premises</th>
<th>Conclusion</th>
<th>Validity</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AII1</td>
<td>All B are A&lt;br&gt;Some C are B&lt;br&gt;Some C are A</td>
<td>Valid</td>
<td>Universal</td>
<td>1</td>
</tr>
<tr>
<td>AII3</td>
<td>All B are A&lt;br&gt;Some B are C&lt;br&gt;Some C are A</td>
<td>Valid</td>
<td>Universal</td>
<td>2</td>
</tr>
<tr>
<td>AII2</td>
<td>All A are B&lt;br&gt;Some C are B&lt;br&gt;Some C are A</td>
<td>Invalid</td>
<td>Universal</td>
<td>1</td>
</tr>
<tr>
<td>AII4</td>
<td>All A are B&lt;br&gt;Some B are C&lt;br&gt;Some C are A</td>
<td>Invalid</td>
<td>Universal</td>
<td>2</td>
</tr>
<tr>
<td>IAI3</td>
<td>Some B are A&lt;br&gt;All B are C&lt;br&gt;Some C are A</td>
<td>Valid</td>
<td>Particular</td>
<td>1</td>
</tr>
<tr>
<td>IAI4</td>
<td>Some A are B&lt;br&gt;All B are C&lt;br&gt;Some C are A</td>
<td>Valid</td>
<td>Particular</td>
<td>2</td>
</tr>
<tr>
<td>IAI1</td>
<td>Some B are A&lt;br&gt;All C are B&lt;br&gt;Some C are A</td>
<td>Invalid</td>
<td>Particular</td>
<td>1</td>
</tr>
<tr>
<td>IAI2</td>
<td>Some A are B&lt;br&gt;All C are B&lt;br&gt;Some C are A</td>
<td>Invalid</td>
<td>Particular</td>
<td>2</td>
</tr>
</tbody>
</table>
Appendix C

Experimental Design Study 1A

<table>
<thead>
<tr>
<th>Valid</th>
<th>Universal</th>
<th>Particular</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBB</td>
<td>BBU</td>
<td>UBU</td>
</tr>
<tr>
<td>AII1</td>
<td>AII3</td>
<td>AII3</td>
</tr>
<tr>
<td>AII1</td>
<td>AII3</td>
<td>IAI3</td>
</tr>
<tr>
<td>AII1</td>
<td>AII3</td>
<td>IAI4</td>
</tr>
<tr>
<td>BBB</td>
<td>BBU</td>
<td>BUU</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>AII4</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>IAI1</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>IAI2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invalid</th>
<th>Universal</th>
<th>Particular</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBB</td>
<td>BBU</td>
<td>UBU</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>AII4</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>IAI1</td>
</tr>
<tr>
<td>AII4</td>
<td>AII2</td>
<td>IAI2</td>
</tr>
<tr>
<td>BBB</td>
<td>BBU</td>
<td>BUU</td>
</tr>
<tr>
<td>AII1</td>
<td>AII2</td>
<td>AII1</td>
</tr>
<tr>
<td>AII1</td>
<td>AII2</td>
<td>IAI1</td>
</tr>
<tr>
<td>AII1</td>
<td>AII2</td>
<td>IAI2</td>
</tr>
</tbody>
</table>
Appendix D

*Study 1A: Semantic Test Syllogisms*

Note: Value in parentheses is the percent believability of the statement based on pre-testing.

<table>
<thead>
<tr>
<th>Study 1A: Semantic Test Syllogisms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AII1</strong> Valid Universal 1</td>
</tr>
<tr>
<td>BBB</td>
</tr>
<tr>
<td>All lizards are reptiles (.75)</td>
</tr>
<tr>
<td>Some pets are lizards (.93)</td>
</tr>
<tr>
<td>Some pets are reptiles (.90)</td>
</tr>
<tr>
<td>BBU</td>
</tr>
<tr>
<td>All denim jeans have two legs (.80)</td>
</tr>
<tr>
<td>Some skirts are denim jeans (.60)</td>
</tr>
<tr>
<td>Some skirts have two legs (.43)</td>
</tr>
<tr>
<td>UBU</td>
</tr>
<tr>
<td>All university courses are free (.25)</td>
</tr>
<tr>
<td>Some expensive courses are university courses (.97)</td>
</tr>
<tr>
<td>Some expensive courses are free (.20)</td>
</tr>
<tr>
<td>AII3 Valid Universal 2</td>
</tr>
<tr>
<td>BBB</td>
</tr>
<tr>
<td>All muffins contain carbohydrates (.80)</td>
</tr>
<tr>
<td>Some muffins are delicious (.89)</td>
</tr>
<tr>
<td>Some delicious things contain carbohydrates (.86)</td>
</tr>
<tr>
<td>BBU</td>
</tr>
<tr>
<td>All grocery lists are reminders (.92)</td>
</tr>
<tr>
<td>Some grocery lists are forgotten (.89)</td>
</tr>
<tr>
<td>Some forgotten things are reminders (.31)</td>
</tr>
<tr>
<td>UBU</td>
</tr>
<tr>
<td>All children are daughters (.29)</td>
</tr>
<tr>
<td>Some children are sons (.92)</td>
</tr>
<tr>
<td>Some sons are daughters (.43)</td>
</tr>
<tr>
<td>IAI3 Valid Particular 1</td>
</tr>
<tr>
<td>BBB</td>
</tr>
<tr>
<td>Some horses are wild (.88)</td>
</tr>
<tr>
<td>All horses are animals (.90)</td>
</tr>
<tr>
<td>Some animals are wild (.96)</td>
</tr>
<tr>
<td>BBU</td>
</tr>
<tr>
<td>Some individuals have foster parents (.88)</td>
</tr>
<tr>
<td>All individuals have birth parents (.97)</td>
</tr>
<tr>
<td>Some birth parents are one's foster parents (.21)</td>
</tr>
<tr>
<td>BUU</td>
</tr>
<tr>
<td>Some strangers are people we ignore (.80)</td>
</tr>
<tr>
<td>All strangers are our best friends (.19)</td>
</tr>
<tr>
<td>Some of our best friends are people we ignore (.45)</td>
</tr>
<tr>
<td>IAI4</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td><strong>Valid</strong>&lt;br&gt;Particular 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BBU</th>
<th>BUU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some cats are hairless (.78)&lt;br&gt;All cats are whiskered (.90)&lt;br&gt;Some whiskered things are hairless (.30)</td>
<td>Some summer days are hot (.95)&lt;br&gt;All hot days are snowy (.23)&lt;br&gt;Some snowy days are summer days (.46)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AII4</th>
<th>BBB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invalid</strong>&lt;br&gt;Universal 1</td>
<td>All bicycles have wheels (.81)&lt;br&gt;Some wheeled objects are modes of transportation (.92)&lt;br&gt;Some modes of transportation are bicycles (.94)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BBU</th>
<th>UBU</th>
</tr>
</thead>
<tbody>
<tr>
<td>All years have months (.97)&lt;br&gt;Some months have 30 days (.98)&lt;br&gt;Some 30 day periods are years (.33)</td>
<td>All triangles are four sided (.24)&lt;br&gt;Some four sided objects are squares (.93)&lt;br&gt;Some squares are triangles (.44)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AII2</th>
<th>BBB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invalid</strong>&lt;br&gt;Universal 2</td>
<td>All hammers are tools (.76)&lt;br&gt;Some handheld objects are tools (.92)&lt;br&gt;Some handheld objects are hammers (.89)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BBU</th>
<th>UBU</th>
</tr>
</thead>
<tbody>
<tr>
<td>All poisonous chemicals are labelled (.85)&lt;br&gt;Some food products are labelled (.95)&lt;br&gt;Some food products are poisonous chemicals (.21)</td>
<td>All surgeons play the violin (.20)&lt;br&gt;Some 5-year-olds play the violin (.89)&lt;br&gt;Some 5-year-olds are surgeons (.13)</td>
</tr>
<tr>
<td>IAI1</td>
<td>BBB</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td><strong>Invalid</strong></td>
<td>Some rules are unfair (.93)</td>
</tr>
<tr>
<td><strong>Particular 1</strong></td>
<td>All team sports have rules (.76)</td>
</tr>
<tr>
<td></td>
<td>Some team sports are unfair (.80)</td>
</tr>
<tr>
<td>BBU</td>
<td>BUU</td>
</tr>
<tr>
<td>Some boats are ships (.94)</td>
<td>Some flowers are blue (.90)</td>
</tr>
<tr>
<td>All canoes are boats (.91)</td>
<td>All trees are flowers (.31)</td>
</tr>
<tr>
<td>Some canoes are ships (.25)</td>
<td>Some trees are blue (.29)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IAI2</th>
<th>BBB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invalid</strong></td>
<td>Some Alaskan animals eat fish (.96)</td>
</tr>
<tr>
<td><strong>Particular 2</strong></td>
<td>All polar bears eat fish (.93)</td>
</tr>
<tr>
<td></td>
<td>Some polar bears are Alaskan animals (.89)</td>
</tr>
<tr>
<td>BBU</td>
<td>BUU</td>
</tr>
<tr>
<td>Some fences are tall (.94)</td>
<td>Some buses are on time (.94)</td>
</tr>
<tr>
<td>All pro basketball players are tall (.89)</td>
<td>All flights are on time (.30)</td>
</tr>
<tr>
<td>Some pro basketball players are fences (.20)</td>
<td>Some flights are buses (.25)</td>
</tr>
</tbody>
</table>
Appendix E

**Study 1A Procedure**

<table>
<thead>
<tr>
<th>Screen 1: introduction to the syllogism</th>
<th>Screen 4: instructions for keyboard usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>The logical syllogism is a form of deductive argument.</td>
<td>Now it’s time for some practice trials!</td>
</tr>
<tr>
<td>It has three parts:</td>
<td>Indicate whether the conclusion is TRUE OR FALSE by pressing:</td>
</tr>
<tr>
<td>The major premise</td>
<td>“z” for TRUE</td>
</tr>
<tr>
<td>The minor premise</td>
<td>or</td>
</tr>
<tr>
<td>The conclusion</td>
<td>“m” for FALSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Screen 2: example of syllogisms</th>
<th>Screen 5: space bar pressed to continue</th>
</tr>
</thead>
<tbody>
<tr>
<td>For example:</td>
<td>If…</td>
</tr>
<tr>
<td>Major: All A are B</td>
<td>[major premise]</td>
</tr>
<tr>
<td>Minor: All C are A</td>
<td>And if…</td>
</tr>
<tr>
<td>Conclusion: Therefore, all C are B</td>
<td>[minor premise]</td>
</tr>
<tr>
<td>All dogs are mammals</td>
<td>Then is...</td>
</tr>
<tr>
<td>All beagles are dogs</td>
<td>[conclusion]</td>
</tr>
<tr>
<td>Therefore, all beagles are mammals</td>
<td>TRUE (‘z’) OR FALSE (‘m’)?</td>
</tr>
</tbody>
</table>

**Screen 3: introduction to the task**

Your job will be to:

Assume the major and minor premises are TRUE.

If they are true, then is the conclusion TRUE or FALSE?

For example:

If... "All A are B"
And If... "All C are A"

Then is the statement "All C are B" TRUE or FALSE?

**Screen 4: instructions for keyboard usage**

Now it’s time for some practice trials!

Indicate whether the conclusion is TRUE OR FALSE by pressing:

“z” for TRUE
or
“m” for FALSE

**Screen 5: space bar pressed to continue**

If…
[major premise]

And if…
[minor premise]

**Screen 6: space bar pressed to continue**

Then is...
[conclusion]
TRUE (‘z’) OR FALSE (‘m’)?

Recall...
[Major]
[Minor]
[Conclusion2]

**Screen 7: validity judgment**

How confident are you that you solved the syllogism correctly?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all confident</td>
<td>Extremely confident</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Italicized material was not viewed by participants.
Appendix F

Study 1B Procedure

<table>
<thead>
<tr>
<th>Screen 1: structure vs. content distinction</th>
<th>Screen 2: explanation of structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a difference between the STRUCTURE and the CONTENT of a syllogism.</td>
<td>STRUCTURE</td>
</tr>
<tr>
<td></td>
<td>- Framework of the syllogistic argument.</td>
</tr>
<tr>
<td></td>
<td>- There are no words, just the terms A, B, and C.</td>
</tr>
<tr>
<td>All A are B</td>
<td>All C are A</td>
</tr>
<tr>
<td>Therefore, All C are B</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Screen 3: how to verify structure</th>
<th>Screen 4: explanation of content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagrams</strong> can illustrate structure:</td>
<td>CONTENT</td>
</tr>
<tr>
<td>All A are B</td>
<td>- Statements within the syllogism.</td>
</tr>
<tr>
<td>All C are A</td>
<td>- A, B, and C are replaced with words.</td>
</tr>
<tr>
<td>Therefore, All C are B</td>
<td>All fish are <strong>animals</strong> All salmon are <strong>fish</strong> Therefore, all salmon are <strong>animals</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Screen 5: how to verify content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evidence</strong> can help us evaluate content:</td>
</tr>
<tr>
<td>All fish are animals → checking the facts:</td>
</tr>
<tr>
<td>Fish belong to the animal kingdom. A fish is not a plant, fungus, or bacterium. The animal kingdom contains:</td>
</tr>
<tr>
<td>Fish</td>
</tr>
<tr>
<td>Reptiles</td>
</tr>
<tr>
<td>Amphibians</td>
</tr>
<tr>
<td>Birds</td>
</tr>
<tr>
<td>Mammals</td>
</tr>
<tr>
<td>Invertebrates</td>
</tr>
</tbody>
</table>

Note: Italicized material was not viewed by participants.
Appendix G

*Experimental Design Study 2*

<table>
<thead>
<tr>
<th>Valid</th>
<th>Easy</th>
<th></th>
<th>Difficult</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congruent</td>
<td>Incongruent</td>
<td>Incongruent + warning signal</td>
<td>Congruent</td>
</tr>
<tr>
<td></td>
<td>BBB</td>
<td>BBU</td>
<td>UBU</td>
<td>BBB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invalid</th>
<th>Easy</th>
<th></th>
<th>Difficult</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congruent</td>
<td>Incongruent</td>
<td>Incongruent + warning signal</td>
<td>Congruent</td>
</tr>
<tr>
<td></td>
<td>BBU</td>
<td>BBB</td>
<td>UBB</td>
<td>BBU</td>
</tr>
</tbody>
</table>
Appendix H

*Study 2: Semantic Test Syllogisms*

Note: Value in parentheses is the percent believability of the statement.

<table>
<thead>
<tr>
<th>AII1</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid Universal 1</td>
<td>All lizards are reptiles (.75)</td>
<td>Some pets are lizards (.93)</td>
<td>Some pets are reptiles (.90)</td>
</tr>
<tr>
<td>All denim jeans have two legs (.80)</td>
<td>Some skirts are denim jeans (.60)</td>
<td>Some skirts have two legs (.43)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AII3</th>
<th>BBB</th>
<th>BBU</th>
<th>UBU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid Universal 2</td>
<td>All muffins contain carbohydrates (.80)</td>
<td>Some muffins are delicious (.89)</td>
<td>Some delicious things contain carbohydrates (.86)</td>
</tr>
<tr>
<td>All grocery lists are reminders (.92)</td>
<td>Some grocery lists are forgotten (.89)</td>
<td>Some forgotten things are reminders (.31)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IAI3</th>
<th>BBB</th>
<th>BBU</th>
<th>BUU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid Particular 1</td>
<td>Some horses are wild (.88)</td>
<td>All horses are animals (.90)</td>
<td>Some animals are wild (.96)</td>
</tr>
<tr>
<td>Some individuals have foster parents (.88)</td>
<td>All individuals have birth parents (.97)</td>
<td>Some birth parents are one’s foster parents (.21)</td>
<td></td>
</tr>
<tr>
<td>Some strangers are people we ignore (.80)</td>
<td>All strangers are our best friends (.19)</td>
<td>Some of our best friends are people we ignore (.45)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix I

Non-Significant Results Pertaining to Believability and Congruency from Studies 1A, 2, and 3

Study 1A

*Reaction Time (ms) on Valid and Invalid Syllogisms for Major Premise x Believability*

<table>
<thead>
<tr>
<th>Major Premise</th>
<th>Believability</th>
<th>Reaction Time (ms)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Valid</td>
<td>Invalid</td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td>BBB</td>
<td>13327.16</td>
<td>19904.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BBU</td>
<td>19736.34</td>
<td>20840.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UBU/BUU</td>
<td>19101.76</td>
<td>20229.85</td>
<td></td>
</tr>
<tr>
<td>Particular</td>
<td>BBB</td>
<td>16021.70</td>
<td>18666.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BBU</td>
<td>24466.60</td>
<td>23636.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UBU/BUU</td>
<td>20417.91</td>
<td>16259.85</td>
<td></td>
</tr>
</tbody>
</table>

Study 2

*Confidence on Valid and Invalid Syllogisms for Congruency x Difficulty*

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Congruency</th>
<th>Confidence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Valid</td>
<td>Invalid</td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td>congruent</td>
<td>4.48</td>
<td>3.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>4.07</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>4.12</td>
<td>3.79</td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>congruent</td>
<td>4.11</td>
<td>3.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>3.80</td>
<td>4.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>4.11</td>
<td>4.01</td>
<td></td>
</tr>
</tbody>
</table>
### Accuracy and Reaction Time (ms) on Invalid Syllogisms for Congruency x Difficulty

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Congruency</th>
<th>Accuracy</th>
<th>Reaction Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>congruent</td>
<td>.81</td>
<td>19772.83</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>.36</td>
<td>18502.39</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>.55</td>
<td>23362.25</td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>congruent</td>
<td>.73</td>
<td>14187.45</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>.19</td>
<td>12770.82</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>.46</td>
<td>16261.99</td>
</tr>
<tr>
<td>Difficult</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Study 3

### Reaction Time (ms) and Confidence on Valid and Invalid Syllogisms for Congruency x Difficulty

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Congruency</th>
<th>Reaction Time (ms)</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Valid</td>
<td>Invalid</td>
</tr>
<tr>
<td></td>
<td>congruent</td>
<td>10223.20</td>
<td>12957.62</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>12912.30</td>
<td>17238.23</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>15547.22</td>
<td>18336.93</td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>congruent</td>
<td>18602.82</td>
<td>20176.25</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>18447.94</td>
<td>17819.35</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>19054.60</td>
<td>22615.07</td>
</tr>
<tr>
<td>Difficult</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Accuracy on Invalid Syllogisms for Difficulty x Congruency

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Congruency</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>congruent</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>.55</td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>congruent</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td>incongruent</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>warning signal</td>
<td>.51</td>
</tr>
<tr>
<td>Difficult</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Belief-content conflict in the premises may even encourage metaphorical and symbolic processing of the syllogism which would otherwise be treated literally. Upon encountering questionable content at the outset of an argument, reasoners may contend with the lack of literal overlap between terms (all sad songs are blue) by looking for subtler symbolic similarities (i.e., the musical blues genre would be emphasized). In this way, confrontation “presses for relation” and so “the common element comes to the fore” (Arnheim, 1969, p. 62), yielding metaphorical understanding that has been driven by an emphasis on the abstract relationships between categories as opposed to the concrete ones (Lakoff & Johnson, 1980). Debatin (1995), writing about the inherent ‘rationality of metaphor’ noted two functions of metaphorical processing that would translate into the syllogistic context. Interpreting a syllogism metaphorically could serve a cognitive creative function by lending meaning to cognitive content that would otherwise be impossible to conceptualize; it could also promote understanding by connecting propositional and explicit with holistic and implicit knowledge.

Encountering an implausible premise may spark in participants a reliance on just this style of thinking because it affords legitimacy to the premises and furnishes a foundation upon which to work with them. Metaphorical thinking entails more than a mere orientation towards the implicit and symbolic qualities of a given category, it also changes the way that parts (categories) are integrated into wholes. For one thing metaphorical thinking may encourage complex relational thinking (Lakoff & Johnson, 1980) in which multiple categories are integrated as opposed to basic relational thinking in which binary relationships are constructed through the comparison of two categories. Metaphorical thinking entails a juggling of many qualities simultaneously; it necessitates the juggling of real as well as surreal qualities of a metaphor’s tenor and vehicle (Richards, 1936). This could help reasoners to integrate the three parts of an argument, facilitating an abstract process of inference-making across the argument as a whole as opposed to a more restrained inference-making between the two domains of a single proposition. Much like hypothetical thinking and the construction of imaginary worlds, metaphorical thinking lead to the liberal treatment of categories; hypothetical thinking constructs a world in which the implausible could be whereas metaphorical thinking emphasizes the implicit and subtle connections underlying otherwise implausible combinations.