Inhibition's Role in Mediating Recall of Arguments

A Thesis Submitted to the
University Honors Program
In Partial Fulfillment of the
Requirements of the Baccalaureate Degree
With Upper Division Honors

Department Of
Psychology

By
James R. Mogan

DeKalb, Illinois

May 12, 2018
University Honors Program
Capstone Approval Page

Capstone Title:
   Inhibition’s Role in Mediating Recall of Arguments

Student Name: James R. Mogan

Faculty Supervisor: M. Anne Britt, PhD

Faculty Approval Signature: [Signature]

Department of: Psychology

Date of Approval: May 3, 2018
AUTHOR: James R. Mogan

THESIS TITLE: Inhibition’s Role in Mediating Recall of Arguments

ADVISOR: M. Anne Britt, PhD

ADVISOR’S DEPARTMENT: Psychology

DISCIPLINE: Cognitive Psychology

YEAR: 2018

PAGE LENGTH: 28

BIBLIOGRAPHY: 3

ILLUSTRATED: n/a

PUBLISHED: No

LIST PUBLICATION: n/a

COPIES AVAILABLE: Hard Copy; Digital Copy at Huskie Commons
Inhibition’s Role in Mediating Recall of Arguments

James R. Mogan

Northern Illinois University
Abstract

Arguments are everywhere and even college students lack the skill to evaluate whether a reason provides minimal support for a claim. Part of this problem may be that when reading an argument, an individual’s personal beliefs about the topic become activated and can then interfere with precise memory for an argument’s claim. Precise representation of the claim must occur long enough for accurate evaluation. The current study assesses whether inhibition skill level correlates with argument recall skill level. Inhibition skill should help the reader maintain a precise representation of the claim while reading the reason and evaluating the claim-reason connection. In the current study, participants read sets of simple claim-reason arguments and evaluated half for agreement and half for quality (i.e. is the claim supported by the reason). Immediately after rating, participants then recalled the claim. Finally, a go/no-go task and letter span task assessed inhibition skill and working memory capacity. It was hypothesized that those skilled at inhibition would score higher on recall than those less skilled at inhibition (while controlling for general working memory skill). This study examines the importance of inhibiting one’s personal beliefs when evaluating and remembering arguments, an important quality in our everyday lives.
Inhibition’s Role in Mediating Recall of Arguments

Argumentation exists in many facets of our everyday lives; from advertisements, to legal proceedings and legislation, to most academic publications. Due to the prevalence of argumentation, extensive research has gone into understanding the evaluation and comprehension of arguments (Britt, Kurby, Dandotkar, & Wolfe 2008; Britt, Kopp, Durik, Blaum, & Hastings, 2016, Dandotkar, Magliano, & Britt, 2016; Hannon & Daneman, 2004; Larson, Britt, & Kurby, 2009; Voss, Fincher-Kiefer, Wiley, & Silfies, 1993, Wolfe & Britt, 2008).

An informal argument has one or more claims supported by one or more reasons to persuade the reader (Britt et al., 2008). The claim is what the author or speaker wants you to believe. In the case of Argument 1, the claim is “Car insurance providers are justified in charging higher rates to those under the age of 25”. The reason is a statement that if accepted by the audience, increases believability in the claim. The reason in Argument 1 is “because younger drivers represent a greater risk for accidents.”

Argument 1. Car insurance providers are justified in charging higher rates to those under the age of 25 because younger drivers represent a greater risk for accidents.

Kuhn (1992) examined what types of individuals evaluated arguments best and what qualities were present in those that did and did not reason well. They found that quality evaluators possessed the ability to reason both sides of an argument and that ability to reason reflected one’s level of education, wherein (as expected) experts reasoned well due to their education and background in argumentation.

In addition to training, evaluation skill may be affected by one’s skill in inhibiting task irrelevant information. When reading an argument, associated information that is activated including beliefs and associated reasons or knowledge (Voss et al., 1993). For some evaluation
tasks like an agreement task (e.g., make a judgment about the extent to which you agree or disagree with the argument), that activated associated information is relevant. But for other evaluation tasks like a quality task (e.g., make a judgment about whether the argument was structurally or logically flawed regardless of whether you agree with it or not), those associated beliefs must be inhibited to accurately complete the task. Prior research has shown that high school and college students have trouble making accurate quality evaluations of simple claim-reason arguments (Larson et al., 2009). Part of the problem of evaluating argument quality may be that when reading the claim and reason, one’s skill in inhibiting related beliefs that interferes with maintaining a precise memory for an argument’s claim. A precise claim is needed to be able to evaluate the claim-reason connection accurately.

The current study assesses whether individuals that are skilled at inhibition are also more skilled at argument evaluation compared to those less skilled at inhibition. Participants read a set of simple claim-reason arguments to evaluate their agreement and another set to evaluate their quality (i.e. whether the reason supports the claim). Immediately after ratings, they were asked to recall the claim. Half of the arguments were unwarranted arguments where the reason did not support the claim and half were warranted arguments where the provided reason could support the claim. To measure individual differences in inhibition skill, a go/no-go task which measures ones’ ability to adopt new rules while inhibiting previous ones (Salas, 2017). I also used the running letter span test as a measure of working memory capacity (Salas, 2017) to be used as a covariate. This will help us understand whether inhibition skill is related to one’s memory for claims, especially for the quality evaluation task that requires inhibiting related beliefs.
In the next section, I will review prior research on evaluating arguments and then I will view working memory and inhibition skills that might influence the evaluation of arguments. Finally, I will present the study hypotheses.

Evaluating arguments

One can evaluate arguments on several dimensions. For example, when given Argument 1, the reader may rate it on agreement (e.g., “rate on a 6-point scale how much you agree or disagree with this argument on a personal level”). One can also rate arguments on its quality (e.g., “make a judgment about whether the argument was structurally or logically flawed regardless of whether you agree with it or not”). In the latter, the task is to decide whether each argument is structurally or logically acceptable or flawed without accounting for one’s belief. For instance, Argument 1 should be rated with at least moderate quality because the reason justifiably supports the claim. Argument 2 should be rated as low quality because the reason fails to support the claim.

Argument 2. Car insurance providers are justified in charging higher rates to those under the age of 25 because younger drivers listen to rap music.

Even college students have trouble evaluating arguments based only on the quality of the argument. Larson et al. (2009) had undergraduates read and evaluate the quality of a set of arguments that varied on their quality. Results found that untrained undergraduates were only slightly better than chance at discriminating structurally good from structurally flawed arguments. In addition, they found that the introduction of a short tutorial significantly improved performance on argument evaluation (Larson et al., 2009).

Evaluating arguments on their quality can be challenging for several reasons. First, the two evaluation tasks target distinct parts of the argument. When given the simple claim-reason
format of arguments, Wolfe and Britt (2008) found that agreement judgements are based more so on one’s agreement with the claim, not the entire argument, whereas quality judgements are based more so on the reason. Therefore, when processing arguments, evaluators emphasize the connection between the reason and the claim (i.e. the warrant) more when evaluating for quality versus agreement. Second, part of the problem is that people do not remember the claim predicate which needs to be attached to the reason (Britt et al., 2008). Britt et al. (2008) explored the inability of students to accurately recall claim predicates (the verb or adjective that emphasizes the goal of the argument). The findings showed that although skilled readers outperformed less-skilled readers, skilled readers only correctly recalled key information about arguments 75 percent of the time (p. 75). They also found that those most skilled in recalling arguments were also better at judging argument quality. Therefore, there exists some key distinction that causes difficulty in both processing and remembering arguments.

Research into argument recall has proposed multiple reasons for this poor recall of claim predicates. One such reason is shallow semantic processing, or replacing flawed information with information that makes sense, thereby making an argument or question more understandable/answerable yet failing to recognize the error contained therein (Barton & Sanford, 1993; Dandotkar et al., 2016; Hannon & Daneman, 2004). For instance, “How many animals did Moses take on the ark? Two” (Hannon & Daneman, 2004, p. 201). Here, the subject substituted Noah for Moses because of semantic overlap (the Bible features both, thereby aiding in confusion). Barton and Sanford (1993) also explored this issue through specific contexts, namely, when posed with the question, “When a bicycle accident/plane crashes, where does one bury the survivors?” Participants correctly spotted the error (i.e., survivors are not
buried) with the bicycle example because there is a weaker association with death and survivors versus the plane crash example (p. 482).

Fuzzy-trace theory (Brainerd & Reyna, 1992) postulates that people rely on gists when remembering because they remain salient longer. One component of Brainerd and Reyna’s (1992) theory surmises that people process information in two parallel ways: 1) verbatim, or exactly as presented and 2) gist, or a simplified, broader interpretation. Of these two ways, the latter remains in the memory for a longer duration. In fact, Britt et al., (2008) did find substitution of argument predicates with a comparable term (i.e. ‘should prohibit’ becomes ‘should ban’ or ‘make illegal’) which supports some role of gist processing.

Another reason for poor recall could be the failure to suppress one’s own belief structure. Voss et al. (1993) found a quick activation of personal beliefs while reading claims. The study had participants judge their agreement with a set of controversial statements twice and then presented them 2 weeks later as arguments. Participants then made either agreement judgments or meaningful judgments. Participants responded as fast to the agreement judgment as to the meaningful judgment. Therefore, it did not take longer to assess one’s attitude toward a statement than it took to understand the statement. This implies that attitudes are activated very quickly while claims are being read. Thus, there may be interference when making evaluations. Working memory is another factor that influences evaluation of arguments, especially argument quality judgments. Dandotkar et al. (2016) found that accurate predicate recallers were also more accurate at making quality judgments, replicating Britt et al. (2008) and demonstrating the importance of working memory with regards to argument comprehension.
Working memory and inhibition

Atkinson’s and Shiffrin’s (1968) multi-store model was extremely successful; however, there were several problems with their ideas concerning the characteristics of short-term memory (i.e. working memory). Baddeley and Hitch (1974) argued that the Multi-Store Model was far too simplistic. According to the Multi-Store Model, working memory holds limited amounts of information briefly with little processing and it is a unitary system (i.e. lacks subsystems). In response to Atkinson and Shiffrin (1968), Baddeley and Hitch (1974) theorized multiple subsystems within working memory (e.g., the visual-spatial sketch pad and phonological loop) moderated by the central executive prior to long-term store. Therefore, working memory is not a single construct. More recently, the processes associated with working memory are being hypothesized (Mikaye, Friedman, Emerson, Witzki, & Howeter, 2000).

The current theory of working memory involves three executive functions: “(1) shifting between tasks or mental sets, (2) updating and monitoring of working memory representations, and (3) inhibition of dominant or prepotent responses” (Mikaye et al., 2000, p. 54). Considering students live in a world of constant distraction with varied concerns, inhibition must play a considerable role in the daily life and study habits of the typical American student with regards to reading comprehension. Prior research into inhibition role in various tasks: Colzato et al., (2008) found a bilingual advantage on reactive inhibition tasks (i.e., inhibition because of more dedicated processing to relevant stimulus). Moreover, Keppel and Underwood (1962) examined inhibition’s role in single-item recall, but the emphasis was not on argument claims.

To the best of my knowledge, only one study has looked at working memory and memory for arguments claims. Mogan, Blaum, & Britt (April, 2017) manipulated the presence of cognitive load (Baddeley & Hitch, 1974) while reading to evaluate arguments for either
agreement or quality. Participants evaluated claim-reason arguments for either agreement or quality task. On half the trials, they had to hold a four-digit number in memory (load condition). Participants then evaluated the argument, recalled it, and recalled the number (for load trials). This study found that the interaction of memory load and task approached significance for claim predicate recall ($p = .063$). The effect of memory load was greater for agreement than quality, replicating the findings of Britt et al. (2008). This suggest that judging agreement is more effected by memory load which is due to the interference of beliefs in the agreement task based on quick activation of beliefs (Voss et al., 1993) and that beliefs are the focus of the agreement task (Wolfe & Britt, 2008). This study demonstrates the importance of inhibiting one’s personal beliefs when evaluating and remembering arguments; especially evaluating argument agreement. The load task was too demanding. Of the 72 participants, only 46 had high enough recall of the numbers to be used in data analysis. Therefore, the current study uses working memory measures (go-no/go and running letter span) as an individual difference measure.

**Current study**

Prior research into inhibition has yet to deal with argumentation. Mogan et al., (April, 2017) had the problem of too many participants being very low on the cognitive load task. Therefore, in the current study, working memory is investigated as an individual difference using a go/no-go task to assess inhibition skill and r-span to assess general working memory (Salas, 2017). The current study assesses whether individuals that are skilled and less skilled at inhibiting are better at argument evaluation which requires inhibition of one’s belief with the expectation that skilled inhibitors will have higher recall accuracy than less skilled inhibitors.

Using the procedure of Britt et al. (2008) and Mogan et al. (April, 2017) as a model, participants rated arguments on how much they agree with it (agreement task) and how
structurally flawed the argument is (quality task). In contrast to the prior studies, evaluation task will be manipulated as a within-subjects factor so that the contrast can be made between tasks. This study will test whether one’s ability to inhibit related information influences argument evaluation and claim predicate recall.

Considering the claim recall measure, it is expected that participants in the current study will replicate Mogan et al., (April, 2017) and Britt et al. (2008), finding that the agreement task will have higher recall accuracy than the quality task and policy claims will have a higher recall than value claims. This is because when reading arguments, readers’ beliefs are activated very quickly (Voss et al., 1993) and those beliefs can interfere with recalling the precise claim; and given the limited capacity of working memory, information involving these interfering beliefs take precedent over the information being asked of the participant. For the agreement task, readers must only focus on their belief regarding the claim, but for the quality task they must focus not only on the claim, but the reason and the claim-reason connection (i.e. warrant) as well (Wolfe & Britt, 2008). Therefore, there will be more potentially distracting information in the quality task than just remembering the claim in detail. Therefore, it is expected that there will be main effects of task and claim type.

This study also predicts both a main effect of inhibition skill and an interaction on recall accuracy. It is predicted that skilled inhibitors will have better recall compared to less skilled inhibitors. This result is expected because skilled inhibitors should be more adept at inhibiting related beliefs and focusing on the argument’s specific assertion. In contrast, less skilled inhibitors will have the most difficulty in recalling the precise claim for the task that requires more inhibiting of related beliefs (quality task).
Method

Participants

43 undergraduate students of Northern Illinois University participated in this experiment as credit for an introductory psychology course.

Design

The design manipulates evaluation task (agreement judgment versus quality judgment) as within-subjects’ factors. See Appendix A for the instructions for each evaluation task. The design also includes a go/no-go task and letter span task to assess inhibition skill and working memory capacity, respectively. The primary dependent measure is recall of the claim predicate.

Materials

Argument Evaluation Task

The experiment used four practice items and forty-four simple arguments, consisting of a claim followed by a reason. Both argument evaluation tasks consisted of 22 arguments each. Two lists were used and counterbalanced so that each argument was presented in each type of task, but a participant received it only in one task. Each set of arguments consisted of 12 warranted (See Argument 1) and 10 unwarranted (See Argument 2) arguments. Additionally, half of arguments utilized a policy claim (i.e. an appeal to change a behavior) and half consisted of value claims (i.e. an appeal to change a belief). To present unique arguments that participants could not familiarize themselves with, there was no repetition of claims or reasons. To equate for recency of mention, one half of the arguments featured the predicate before the theme, whereas the other half had the theme before the predicate. Refer to Appendix B for the arguments presented in this study.
**Go/no-go Task**

The go/no-go task tests one's ability to not respond to a previously learned response (Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003; Salas, 2017). It consists of three color tiles (blue, green, and white) and the participant must only click the mouse when the green-colored tile appears. This is the “Go” stimulus which is presented at a high frequency (80% of trials) versus the blue tile, the “no-go” stimulus, presented at low frequency (20% of trials). The white tile serves as an intermediary between the green and blue. The participants start each trial with a point to attend to (500 ms) followed by the stimulus (100 ms) and then a blank screen (i.e. the white tile). Participants must respond as quickly as possible. A feedback display features the time (in milliseconds) after each correct click of the mouse and serves to encourage quick response. Moreover, participants are given feedback about trials that take too long or are incorrect. This task consists of 300 trials and takes approximately 10 minutes to perform.

**Running Letter Span Task**

The running span (Broadway and Engle, 2010; Salas, 2017) measures participant’s working memory capacity. Participants were presented with a series of letters and were instructed to recall the last letters in the series, ranging from 3 to 7. For each trial, participants saw between 4 and 9 letters three times each. Each letter was displayed for 500 ms successively with set length randomized for each participant. The stimulus lasted 300 ms and the inter-stimulus was 200 ms. After the end of each set, the participants made their responses by clicking the cells of a 3 X 4 that featured all letters from the set. The response screen reminded participants the number of letters intended for recall. The task took approximately 10 minutes to complete.
Procedure

Participants completed the entire experiment in individual rooms using the computer software, E-prime (Psychology Software Tools, Pittsburgh, PA). Participants were randomly assigned to list and task order so that all arguments were presented in both type of tasks and so that half of the participants did the agreement task first and half did the quality task first. In both the agreement and quality conditions, participants begin with instructions and 2 practice items with feedback for task familiarization. The general instructions told participants that they would read several short informal arguments one phrase at a time. The first phrase disappeared after advancing so they were not able to read the argument. They were asked to read the arguments at a normal rate of speed and not think too much about them. For the agreement task, participants were asked to “make a judgment about the extent to which you agree or disagree with the argument” on a 6-point scale with 1 meaning “very strongly disagree” and 6 meaning “very strongly agree”. For the quality task, participants were asked to “make a judgment about whether the argument was structurally or logically flawed regardless of whether you agree with it or not. You want to consider only the structure of the argument not how convincing it is given everything you know.” For this 6-point scale, a 1 meant the argument is “structurally good or not flawed” and a 1 meant it is “extremely flawed”.

For each argument, participants press the spacebar to present the claim for reading. Then they pressed the spacebar to remove the claim and present the reason for reading. Next, they pressed the spacebar to remove the reason and to the agreement or quality evaluation task. Subsequently, the participants recalled the claim of the argument by typing it in a textbox. For this recall task, participants were asked to recall the argument using “as close to the exact words”
as possible. Then they completed the go/no-go task followed by the running span task. Finally, they were thanked and debriefed.

**Analyses**

Recall was scored for verbatim claim predicate recall (i.e. “has the right to intervene” was counted incorrectly if changed to “is right” or “should”) by myself and my advisor with a kappa of 0.98. A third rater was then brought in to argue disagreements. Accuracy for the go/no-go task was computed within E-prime as the number of correct responses out of the total trials. Descriptive statistics were calculated to make sure that performance is within the expected range and assumptions tested. First a 2 Evaluation task (agreement versus quality) X 2 Claim type (policy versus value) repeated measures ANOVA on claim predicate recall was conducted to ascertain a replication of Britt et al. (2008) and policy advantage in recall. Second, a 2 Evaluation task (agreement versus quality) X 2 Warrant type (flawed versus acceptable) repeated measures ANOVA on claim predicate recall was conducted with performance on the Go-No-go task as a covariate. There were not enough participants to analyze the effect of inhibition skill, so it was only used as a covariate. Additionally, we conducted these same analyses on response time.

**Results**

A 2-evaluation task (agreement versus quality) X 2-claim type (policy versus value) repeated measures ANOVA on claim predicate recall found a significant main effect of claim type, $F(1, 42) = 22.78$, MSE = .021, $p < 0.05$. As shown Figure 1, the results replicated Britt et al. (2008) in that policy claims were recalled more than value claims. No other results were significant. A 2-evaluation task (agreement versus quality) X 2-warrant type (flawed versus acceptable) repeated measured ANOVA on claim predicate recall found a significant main effect
of warrant type, $F(1, 41) = 21.88$, MSE = .013, $p < 0.05$ and a significant interaction of task by warrant type, $F(1, 41) = 4.19$, MSE = .017, $p < 0.05$. As shown in Figure 2, flawed arguments exhibited higher recall than acceptable arguments and recall was significantly higher for flawed arguments in the agreement evaluation than in the quality evaluation, but the difference between the two warrant types was not significant for acceptable arguments in either evaluation task. The same analysis also yielded a significant interaction of warrant type and performance on the go/no-go task, $F(1, 41) = 4.18$, MSE = .017, $p < 0.05$ and an interaction approaching significance for claim type and performance on the go/no-go task, $F(1, 41) 3.85$, MSE = .017, $p = 0.057$, meaning with an increase in inhibitory skill, there was an increase in one’s ability to accurately recall claim predicates. Lastly, the inclusion of rating response time (Figure 3) as a dependent variable also yielded a significant interaction of task and warrant type, $F(1, 41) = 4.33$, MSE = 1.63, $p < 0.05$. For flawed arguments, participants responded more slowly in the agreement task compared to the quality task, however, there was no difference on response times between tasks for good arguments.

**Discussion**

The primary goal of this study was to test how inhibition skill influences one’s ability to evaluate and recall argument claim predicates. Unfortunately, due to an insufficient sample size and performance on the go/no-go task being near ceiling, the study could not accomplish this goal. Still, inhibitory skill did function as a covariate and significantly correlate with one’s ability to recall arguments with increasing inhibitory skill leading to increased argument claim predicate recall.

Additionally, the current study found policy claims exhibited higher recall than value claims, which coincides with past literature examining argument recall (Britt et al, 2008;
Dandotkar et al., 2016; Mogan et al., 2017). Interestingly, when comparing Mogan et al. (2017) with the current study, the advantage of policy claim recall holds true regardless of task condition and regardless of the measure being a within- or between-subjects; making these findings incredibly generalizable.

Moreover, this study also discovered that the warrant significantly impacts one’s ability to recall the predicate of a claim, with an advantage for flawed arguments being recalled more than acceptable arguments, especially in the agreement task over the quality task. This advantage for flawed arguments also holds true regardless of one’s inhibitory skill (i.e. one’s performance on the go/no-go task); however, argument evaluation skill does increase as inhibition skill increase as well in line with our hypothesis. Additionally, better performance on the go/no-go task also increased with one’s ability to recall argument predicates. Lastly, participants rated arguments slower in the agreement task compared to the quality task, however, this finding only applied to the agreement task and the opposite was true for the quality task.

With regards to rating one’s agreement to an argument, the results indicated that there is no difference between tasks when rating for acceptable arguments, yet interestingly, when the argument is flawed, rating for quality is faster than rating for agreement. This suggests that there is minimal processing in the quality condition if the argument is blatantly wrong, but when evaluating for agreement, one must weigh their personal beliefs against not only the claim, but all the components of a simple argument (i.e. the claim, reason, and warrant). These findings demonstrate the need for future research to include an inventory or assessment of one’s personal beliefs, a key component has previously been ignored (Voss et al., 1993; Mogan et al, 2017).
Moreover, these findings are in line with Rouet, Bigot, de Pereyra, and Britt (2016), which demonstrated that discrepancies cause information (in this case, a simple argument) to stand out and therefore be more likely to be recalled.

The largest take away from this study that also speaks to future research would be the clear distinction between the two argument evaluation tasks. With the agreement task, the scrutiny of the argument varies with a change in the warrant (i.e., from good to bad). Specifically, if an argument is bad, rating response time increases, suggesting more scrutiny is involved, which leads to higher recall. If the argument is acceptable, there is no need for scrutiny and ergo, less time required for recall, but recall worsens. For quality evaluation, one must scrutinize the argument regardless of one’s agreement with the argument, so when an argument is acceptable, rating response time and recall of the argument claim predicate are comparable to the agreement task because both require scrutiny. However, when an argument is clearly bad, evaluation for quality becomes less mentally taxing and ceases to need further scrutiny the argument, which explains the decreased rating response time.

Limitations

There are a couple of limitations of this study. First, given the lack of participants, more subjects are needed to truly assess the impact that inhibitory skill has on argument claim predicate recall. Second, the experimented tested the executive functions of inhibition and capacity, but failed to test the third executive function, shifting. To correctly assess how executive functions factor into argument comprehension and recall, the inclusion of an additional test for shifting (e.g., number-letter, plus-minus) is necessary. Finally, the population could be widened to included people outside the NIU SONA subject pool so that the results are more generalizable.
Future Direction

There are a couple future directions that may be interesting. Clearly, given the extremely different outcomes between the agreement task and the quality task, more research is necessary to understand the mechanisms that lead to the two different outcomes seen in this experiment. Also, 4 participants were lost due to the E-prime program closing part-way through the experiment, which at the time, had been understood to be user error (i.e. the participant pressed a key that forced the program to end); however, the only way to cause this would be pressing the Windows key. Future research needs to account for this potential shortcoming by disabling this key. Lastly, a measure to assess one’s personal beliefs could elucidate how said beliefs impact argument claim predicate recall and potentially explain the distinct outcomes between evaluation tasks.
References


Appendix A

Agreement Instructions

During this set of arguments, you will be asked to make a judgment about the extent to which you agree or disagree with the argument. Select 6 if you “very strongly agree”, 5 if “strongly agree”, 4 if you “somewhat agree”, 3 if you “somewhat disagree”, 2 if you “strongly disagree” and 1 if you “very strongly disagree”.

After making your judgment, you will be asked to recall the first phrase of the argument using as close to the exact words as possible when reading as you normally would read an argument.

Please do not pause during the reading of the argument.

Please take breaks when you need them but only take a break after your judgement and between arguments. Please do not pause during the reading or evaluation of an argument.

Quality Instructions

During this set of arguments, you will be asked to make a judgment about whether the argument was structurally or logically flawed regardless of whether you agree with it or not. You want to consider only the structure of the argument not how convincing it is given everything you know. A 6 means the argument is “structurally good or not flawed” and a 1 means it is “extremely flawed”. Select 6 if you “extremely good structure”, 5 if “very good structure”, 4 if you “somewhat good structure”, 3 if you “somewhat flawed structure”, 2 if you “very flawed structure” and 1 if you “extremely flawed structure”.

After making your judgment, you will be asked to recall the first phrase of the argument using as close to the exact words as possible when reading as you normally would read an argument. Please do not pause during the reading of the argument.

Please take breaks when you need them but only take a break after your judgement and between arguments. Please do not pause during the reading or evaluation of an argument.
Appendix B

Arguments in blocks. List 1 was blocks 1 and 2 for agreement evaluation and blocks 3 and 4 for quality evaluation. List 2 consists of the same block sequence with the argument evaluation tasks switched.

<table>
<thead>
<tr>
<th>Block 1</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We have wasted billions of dollars fighting a war on drugs because we do not go after those who launder the money.</td>
<td>acceptable</td>
<td>After</td>
<td>Value</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Americans should support a two-party system because the two parties are the republicans and democrats.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The U.S. is right to intervene in other countries’ affairs because local events can catastrophically impact the entire world.</td>
<td>acceptable</td>
<td>After</td>
<td>Value</td>
<td>flawed</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rollercoasters are dangerous because they are large and metal.</td>
<td>flawed</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimentation on animals is a useful research activity because it has led to crucial medical advances.</td>
<td>acceptable</td>
<td>Before</td>
<td>Value</td>
<td>acceptable</td>
<td>Before</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex education should be taught in public high-schools because it will improve students’ decisions at a critical time in their lives.</td>
<td>acceptable</td>
<td>Before</td>
<td>Policy</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prayer should be forbidden in public schools because it doesn't help you get into college.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloning should not be legal in our society because it will be used for large scale organ harvesting.</td>
<td>acceptable</td>
<td>Before</td>
<td>Policy</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People should be permitted to have only two biological children because children are small.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We should require forest fires to burn in a controlled way because dead underbrush puts the whole area at risk.</td>
<td>acceptable</td>
<td>After</td>
<td>Policy</td>
<td>flawed</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician assisted suicide for terminally ill patients is advantageous for society because it is legal in the United States.</td>
<td>flawed</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block 2</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is a valuable experience for students to travel the world because it will help students gain deeply meaningful knowledge.</td>
<td>acceptable</td>
<td>After</td>
<td>Value</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
</tr>
<tr>
<td>Recreational use of marijuana should be decriminalized because it does not have any side effects.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Kids in school should not be forced to recite the Pledge of Allegiance because not all students speak English as their first language.

A human expedition to mars would be worthwhile because millions of dollars can be diverted from other space projects.

Small towns should discourage large chain stores from relocating near them because they often eliminate local businesses in small towns.

Owning a pet should be required by law because research shows it prolongs life expectancy.

Legalizing river boat gambling would benefit cities because it provides a huge source of revenue from tourism and taxes.

Handguns decrease crime because handguns are constitutionally protected by the second amendment.

Recycling should be federally mandated because it helps to protect the environment.

Car insurance companies are justified in charging higher rates to those under the age of 25 because younger drivers represent a greater risk of accidents.

People should drive SUVs because they are safer in an accident.

Employers should provide their workers with annual vacations because people need time to spend with their families.

People should limit their garden to plants native to the region because this will most clearly reflect the natural beauty of the area.

Banks shouldn't charge ATM fees because banks are financial institutions.

It is improper for companies to use sweatshops because they often mistreat workers in poor countries. The government should do more to reduce smoking in public because not all people are smokers.

Pornography should be banned in our society because it promotes violence toward women.

It is a recommended idea to promote modern families eating dinner together at the kitchen table because extended families share one roof.
Hunting for sport is helpful for society because otherwise some animal species would overpopulate the earth.
Selling soda and candy bars in high schools is unhealthy for students because most kids do not have much money.
Banning cell phone use while driving is unfair because people are on their cell phones all the time.
Natural foods are not always healthier than processed foods because many times they are actually higher in sodium and fat.

<table>
<thead>
<tr>
<th>Block 4</th>
<th>Type</th>
<th>Location</th>
<th>claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>The state highway speed limit should remain 65 miles per hour because it balances safety with convenience.</td>
<td>acceptable</td>
<td>Before</td>
<td>Policy</td>
</tr>
<tr>
<td>It is wrong to burn an American flag because the government should punish those who don’t support freedom.</td>
<td>flawed</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>It is important that children with violent temperaments be given treatment at a young age because it is easier to cure such behavior in younger children.</td>
<td>acceptable</td>
<td>After</td>
<td>Value</td>
</tr>
<tr>
<td>Attending graduate school should be encouraged for most students because it enables them to find a better job.</td>
<td>acceptable</td>
<td>Before</td>
<td>Policy</td>
</tr>
<tr>
<td>Volunteering is rewarding to students because the government can't be responsible for doing everything.</td>
<td>flawed</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>Libraries should not be allowed to use filters on their internet software because they could screen out critical information.</td>
<td>acceptable</td>
<td>After</td>
<td>Policy</td>
</tr>
<tr>
<td>Vitamins are not harmful because they are necessary for healthy development and living.</td>
<td>acceptable</td>
<td>Before</td>
<td>Value</td>
</tr>
<tr>
<td>Curfews should be placed on teenagers because nighttime is the only time teens ever break the law.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
</tr>
<tr>
<td>The state's minimum wage should be raised because other states have lower taxes.</td>
<td>flawed</td>
<td>Policy</td>
<td></td>
</tr>
<tr>
<td>It is ethical for universities to randomly test athletes for steroids because they provide unfair muscle growth to their users.</td>
<td>acceptable</td>
<td>After</td>
<td>Value</td>
</tr>
<tr>
<td>Companies are overcharging for gas because companies are becoming more global.</td>
<td>flawed</td>
<td>Value</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Figure 1: Predicate Recall by Claim Type

Figure 2: Predicate Recall by Warrant Type
Figure 3: Rating Response Time by Warrant Type

![Graph showing the rating response time by warrant type. The x-axis represents the warrant type (Flawed and Good), and the y-axis represents the response time (s). The graph includes two lines, one for Agree and one for Quality.]