Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hdsp20

Processing and Representation of Arguments in One-Sided Texts About Disputed Topics

Michael B. Wolfe [a], Shawna M. Tanner [a] & Andrew R. Taylor [a]

[a] Psychology Department, Grand Valley State University

Accepted author version posted online: 08 Aug 2013.Published online: 01 Oct 2013.

To cite this article: Michael B. Wolfe, Shawna M. Tanner & Andrew R. Taylor (2013) Processing and Representation of Arguments in One-Sided Texts About Disputed Topics, Discourse Processes, 50:7, 457-497, DOI: 10.1080/0163853X.2013.828480

To link to this article: http://dx.doi.org/10.1080/0163853X.2013.828480

Discourse Processes, 50:457–497, 2013

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/0163853X.2013.828480

Routledge
Taylor & Francis Group

Processing and Representation of Arguments in One-Sided Texts About Disputed Topics

Michael B. Wolfe, Shawna M. Tanner, and Andrew R. Taylor
Psychology Department
Grand Valley State University

We examine students' processing and representation of arguments and counter-arguments in one-sided scientific texts. In Experiment 1, students read texts about evolution and TV violence. Sentence reading times indicated that subjects slowed down reading to the extent that arguments were both more consistent, and inconsistent, with the text position. We refer to this processing pattern as *argument-focused processing*. We also examined whether students hold their beliefs for evidence- or affect-based reasons (belief basis). For the evolution texts, belief basis moderated argument-focused processing. In Experiment 2, subjects read a one-sided text, then a neutral text, and then wrote a summary of the neutral text. Compared to affect-based subjects, evidence-based subjects wrote summaries that were more neutral. Beliefs predicted few differences in processing or representation. We conclude that subjects engage in argument-focused processing when reading one-sided scientific texts. We tentatively conclude that argument-focused processing is moderated by belief basis, but not subject beliefs.

INTRODUCTION

Students frequently need to comprehend information about disputed scientific topics. Examples include evolution, the causes of obesity, the effects of television (TV) violence, and global warming. Often, texts about disputed topics are presented as one-sided arguments that are designed to support a particular

Correspondence concerning this article should be addressed to Michael B. Wolfe, 2224 ASH, Psychology Department, Grand Valley State University, Allendale, MI 49401, USA. E-mail: wolfem@gvsu.edu

position on a topic. In some cases one-sided argumentative texts present counterarguments, typically for the purpose of refuting or discounting the counterarguments (Wolfe, Britt, & Butler, 2009). Previous research suggests that students have a hard time comprehending the arguments and counterarguments in these one-sided texts (Wolfe, Britt, Petrovic, Albrecht, & Kopp, 2009). For example, the reading portion of the grade 12 National Assessment of Educational Progress (National Center for Education Statistics, 2010) includes evaluation of arguments and counterarguments in informational text as one of the assessed skills. Only 5% of students scored at the advanced level, suggesting that many students are not skilled in recognizing or evaluating different arguments in a one-sided argumentative text. In this research, we address students' argument comprehension by examining the extent to which students differentially process and mentally represent arguments and counterarguments in extended one-sided argumentative texts. We refer to processing that is specific to the arguments in these types of texts as *argument-focused processing.*

In addition to characterizing students' argument-focused processing, we examine two individual difference variables as potential moderators of argument-focused processing. First, we examine students' prior beliefs about these topics. We define a belief in this context as a statement of the truth value of a proposition about which evidence could be collected that would potentially support or refute the proposition. Examples include the belief that genetics determine obesity or that watching TV violence causes real violence. Second, we are interested in variations in students' claims about *why* they hold their beliefs (belief basis). In particular, students may claim to hold their beliefs because of evidence-based or affect-based reasons (Griffin, 2008). For both beliefs and belief basis, we also examine whether different processing patterns result in different mental representations of the information.

One-Sided Texts About Disputed Topics

One-sided argumentative texts are characterized in Figure 1 in terms of the relationships between text position, text sentences, and reader beliefs. Consistent with Kintsch's (1998) comprehension theory, the text position is part of the macrostructure. An example from the current experiments is a text that supports the position that watching TV violence causes real violence (referred to as the "TV Yes" text). In these texts, macrostructure cues are stated clearly in both the title and the first paragraph of the text. The reader should understand that the subsequent evidence and arguments are designed to support the text position.

The relationship between the text and the reader can be characterized at two different levels of specificity. At the *sentence level*, each sentence varies in the extent to which it provides support for the text position. Kintsch (1998) refers to

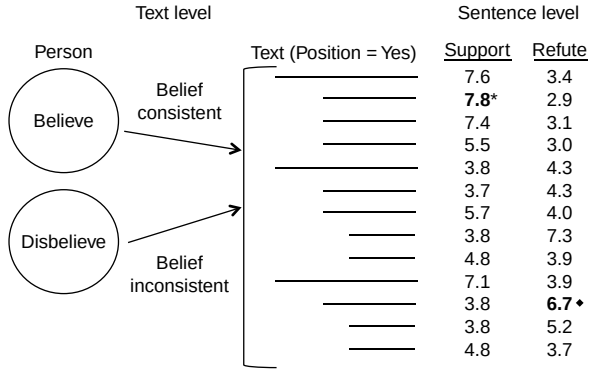

FIGURE 1 The relationships between a reader and a one-sided argumentative text. Horizontal lines represent the sentences of a one-sided text that presents evidence and arguments supporting a proposition. Subjects will believe or disbelieve this proposition and, as a result, their beliefs will be consistent or inconsistent with the text position. Support and refute ratings represent hypothetical ratings for individual sentences. *A sentence that highly supports the proposition. ◆A sentence that highly refutes the proposition. Based on support and refute ratings, individual sentences can be consistent or inconsistent with either the position of the text or subject beliefs.

this level as the microstructure. Consider the following sentences from the TV Yes text:

(1.1) "The more violent of these felons were the most likely to report having learned techniques from television."
(1.2) "In Television and Aggression (1982), Milavsky and his associates reported that television violence had no effect upon children's behavior."

We refer to sentence (1.1) as a *position consistent* sentence, which is a sentence that presents evidence or reasons that support the text position. Sentence (1.2) is a *position inconsistent* sentence.[1] Within the context of the text, sentence (1.2) is presented as evidence supporting a counterargument, namely that watching TV violence does not promote real violence. Each sentence can be characterized in terms of the extent to which it both supports and refutes the text position. Support and refute ratings are based on subject judgments and are on a continuous scale. On average, a Yes text contains sentences that are higher in support than refute ratings. But the subset of sentences that are relatively position inconsistent tends to be rated as more refuting than supporting. At the *text level*, each text either

[1]We consider the terms *position inconsistent* and *counterargument* to mean the same thing. We use *position inconsistent* throughout the manuscript to facilitate the contrast with *belief inconsistent* sentences and texts.

supports or refutes the proposition. Reader beliefs are then crossed with the text position (the reader could also be neutral, which we do not address here). This design allows us to examine circumstances in which readers are comprehending texts that are either belief consistent or belief inconsistent.

We interpret the processing and representation of these texts consistent with Kintsch's (1998; van Dijk & Kintsch, 1983) comprehension theory. According to the theory, texts are mentally represented at multiple levels. The *textbase* is a mental representation of the concepts in the text organized consistent with the text structure. The *situation model* is a mental model that includes the basic knowledge of what the text says, inferences generated during or after comprehension, any opinions or affective responses to the content, and relevant prior knowledge that has been activated and added to the mental model. When a reader constructs a representation of a text that is relatively faithful to the text macrostructure, then the textbase and situation model representations are similar. If, however, processing effort does not correspond to the macrostructure, then the reader will form a situation model that differs from what would be predicted based on the macrostructure.

Argument-Focused Processing of One-Sided Texts

Previous research does not provide clear evidence as to whether students specifically attend to position consistent and inconsistent information during argument comprehension. Wolfe, Britt, Petrovic, et al. (2009) gave college students short one-sided texts that included position inconsistent arguments. Subjects attempted to identify the position inconsistent theme in these arguments but were successful for less than half the texts. In other studies, students generate arguments designed to support one side of an issue. For example, Wolfe and Britt (2008) gave subjects the opportunity to study arguments on both sides of a controversial issue. Students then generated an argument for a side that was assigned to them. In the study phase, subjects tended to view most or all arguments related to both sides. In the argument construction task, however, many subjects included only arguments on the side they were assigned to defend. Wolfe and Britt concluded that many subjects use an argument schema in which a good argument consists of only arguments that support your side. This *myside bias* was specific to the side of the issue subjects were assigned, not to their prior beliefs about the topic. In another argument generation task, Kuhn, Shaw, and Felton (1997) assessed adolescents' and adults' inclusion of information related to both sides of a controversial issue. In this study, only 30% to 35% of both adolescents and adults included any information on the other side of the argument they generated.

However, some evidence does suggest that subjects are sensitive to both position consistent and inconsistent arguments. Wolfe, Britt, and Butler (2009) had subjects read short argumentative texts that included or did not include

position inconsistent arguments. Subjects' judgments of the quality of arguments was higher when the arguments presented and rebutted position inconsistent arguments. In a meta-analysis, Allen (1991) found two-sided arguments with refutations of the position inconsistent arguments to be more persuasive than one-sided arguments. These results suggest a difference between processing position inconsistent arguments on the one hand and specifically identifying or using them on the other.

Three questions are not addressed in this literature that we wish to address in the current studies. First, do readers track the components of arguments as they read? Previous research has not included detailed processing analyses of position consistent and inconsistent arguments. Based on data suggesting subjects prefer two-sided arguments to one-sided arguments, we predict that subjects will discernibly track both position consistent and position inconsistent arguments as they read. We measure argument-focused processing by regressing sentence-by-sentence reading times on the support and refute ratings for each sentence. Evidence for argument-focused processing is obtained if subjects slow down reading as a function of position consistent or inconsistent argument strength, after factoring out reading time variance due to standard comprehension factors. If we obtain evidence for argument-focused processing, two potential processing foci may arise. In a *position consistent processing* focus, processing effort emphasizes the position consistent text information. If readers process argument information consistent with a myside bias, we expect that readers will selectively slow down reading for arguments that support the text position. Alternatively, readers may selectively slow down reading more for position inconsistent information. In that case, readers may be adopting a strategy we refer to as *balanced evidence processing*, which we interpret as an effort to create a situation model of the topic that is more balanced with respect to evidence and arguments than the one-sided macrostructure of the text would suggest.

The other questions we address relate to potential individual difference variables that may moderate argument-focused processing. The second question is whether argument-focused processing changes as a function of subjects' beliefs or the reason why they hold their beliefs. Third, do differences in argument-focused processing result in different situation model representations? In particular, do individual differences in processing result in differences in the extent to which subjects' situation model representations deviate from the text macrostructure? These second two goals are elaborated in the next section.

Beliefs and Belief Basis as Potential Moderators of Argument-Focused Processing

We are interested in the potential influence of beliefs on processing and representation for two reasons. First, students are sometimes called on to

comprehend scientific information that is consistent with or contrary to their beliefs. Second, despite the fairly large body of literature on some aspects of beliefs, such as belief change (e.g., Dole & Sinatra, 1998; Murphy & Mason, 2006), relatively little research addresses students' processing and representation of argumentative texts as a function of belief in the arguments. Wolfe and Britt (2008) found that belief in arguments did not predict bias in terms of which arguments subjects chose to read. Beliefs also did not predict which arguments subjects used in their essays. In a study using connected text, Maier and Richter (2013) had subjects read multiple conflicting texts that were either belief consistent or inconsistent. Sentence recognition data suggested that memory for textbase content was stronger for belief inconsistent texts, but only when a belief inconsistent text was presented first. Memory for situation model content was stronger for belief consistent texts, but only when the texts were presented in a block-by-block format rather than an alternating format. Maier and Richter argue (consistent with Wiley, 2005) that beliefs serve as schemas that can potentially drive processing and representation based on whether information is schema consistent or inconsistent. Britt, Kurby, Dandotkar, and Wolfe (2008) had subjects read and recall simple one-sentence arguments, then state whether they agreed with them or not. Agreement did not predict any overall memory differences. Overall, results in the argumentation literature do not appear to lead to a clear prediction about potential belief effects. As far as we are aware, no research directly addresses belief influences on processing and representation of extended one-sided argumentative texts, as illustrated in Figure 1.

In another body of research related to beliefs, researchers in social psychology examine memory for arguments on both sides of issues relating to attitudes such as abortion or capital punishment (Eagly, Chen, Chaiken, & Shaw-Barnes, 1999; Eagly, Kulsea, Brannon, Shaw, & Hutson-Comeaux, 2000; Roberts, 1985). Subjects in these studies typically read short arguments that are consistent or inconsistent with their attitudes. After reading, subjects recall as many arguments as they can on both sides of the issue. Results of experiments and meta-analyses indicate that subjects do tend to recall attitude consistent arguments better than attitude inconsistent arguments. The effect is small and inconsistent, however, and several studies show no attitude consistent advantages (Greenwald & Sakumura, 1967). To explain the lack of attitude consistent advantages, Eagly et al. (2000) argue that subjects put extra processing resources into attitude inconsistent information in an effort to build up a mental "case" against the attitude consistent position. This motivated processing hypothesis is known as *defense against attack.* Evidence supporting this hypothesis in the current study will be found if subjects slow down reading systematically for belief inconsistent arguments.

We are also interested in potential processing and situation model differences as a function of the reason readers hold their beliefs (belief basis). Belief basis is a

relatively new and under-studied construct. Griffin (2008; Griffin & Ohlsson, 2001) draws a distinction between two types of belief basis. A student who is *evidence-based* for a topic claims that his or her belief about that topic stems from consistency with scientific evidence. In contrast, an *affect-based* student for a topic claims to hold beliefs about that topic for emotional reasons. According to this distinction, students differ in the extent to which beliefs are conceptually tied to evidence that supports those beliefs. Evidence-based students claim their beliefs are tied to scientific evidence. Affect-based students may or may not have any conceptual connection between beliefs about a topic and scientific evidence. Belief basis is construed and measured separately for each topic within each person (see Appendix A). As a result, when discussing belief basis for specific topics, we refer to subjects as evidence- or affect-based for that topic.

In psychometric work, Griffin (2008) conducted a principle components analysis to assess the coherence of the evidence- and affect-based constructs for each of eight topics. Seventy percent to 78% of the variance across the five questions was accounted for with two dimensions, corresponding to the evidence-based and affect-based questions. In work related to belief basis and comprehension (Griffin, 2004), subjects reported their belief and belief basis with respect to evolution and race-based differences in intelligence. Subjects read one-sided argumentative texts and then completed a series of comprehension tests. Evidence-based subjects performed better than affect-based subjects on some of the comprehension measures. Griffin concluded that regardless of whether the evidence is consistent or inconsistent with the subjects' beliefs, evidence-based subjects for a topic are more successful at integrating new evidence related to the topic. In the current research, we go beyond Griffin's work by examining the extent to which belief basis accounts for differences in argument-focused processing and representation.

Beliefs and belief basis are examined in both experiments. In Experiment 1, we examine argument-focused processing and memory differences as a function of both belief consistency and belief basis. In Experiment 2, we examine differences in situation model representations as a function of belief consistency and belief basis. If subjects put more processing emphasis on belief consistent or inconsistent arguments, then situation model representations should deviate from the text macrostructure in a way that is consistent with their processing emphasis. For belief basis, we examine whether subjects differ in position consistent processing or balanced evidence processing as a function of belief basis. For example, evidence-based TV violence subjects may engage in more balanced evidence processing in an effort to understand evidence and arguments related to both sides of the TV violence issue (Griffin, 2008). If so, we expect evidence-based TV violence subjects to create situation models that are more balanced with respect to evidence than the situation models of affect-based TV violence subjects.

EXPERIMENT 1

In Experiment 1, subjects completed a prescreening test to determine their beliefs and belief basis for two topics, evolution and the extent to which TV violence causes real violence. Subjects then read a one-sided text containing evidence designed to support the proposition for each topic as being either true or false (e.g., that all species did or did not evolve from a common ancestor). Subject beliefs were crossed with text position such that approximately half the subjects read belief consistent texts and half read belief inconsistent texts. Finally, subjects completed a memory task in which they recalled as much as possible from a section of each text they read. The cued-section recall task assesses both recall of content that was read and intrusions from other sections of the text and prior knowledge. Intrusions of text content from other text sections can be an indication that subjects have integrated text content into a situation model that does not precisely follow the text macrostructure (Mannes & Kintsch, 1987).

Sentence reading times were analyzed by coding each sentence on eight dimensions. Each dimension corresponds to a psychological factor that may influence reading time. A relative change in reading times as a function of a particular dimension suggests that the subject is mentally tracking that dimension during reading (Graesser, Hoffman, & Clark, 1980). Six dimensions were taken from prior research and were chosen to assess processing of "traditional" (i.e., well established) textbase and situation model representations (syllables, new argument nouns, argument overlap, topic sentences, unitizing sentences, and serial position of a sentence within a section.) The final two dimensions (support and refute) are new and designed to assess argument-focused processing across the sentences of each text.

Textbase and Situation Model Dimensions

The number of *syllables* per sentence is a word-level variable (Stine-Morrow, Miller, Gagne, & Hertzog, 2008; Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998). The two textbase variables are the number of *new argument nouns* and *argument overlap* (Graesser et al., 1980; Stine-Morrow et al., 2008; Zwaan et al., 1995, 1998). New argument nouns indicate the number of noun concepts in each sentence that appear in the text for the first time. Argument overlap is a binary variable that codes whether a sentence shares a noun concept with the immediately preceding sentence. Readers who put effort into textbase construction should slow down as the number of new argument nouns increases and speed up for sentences that have argument overlap compared with those that do not. Three situation model variables taken from Britton (1994; also Stine-Morrow et al., 2008) are the presence of *topic sentences*, *unitizing sentences*, and the *serial position within a section* of the text. The topic sentence code

indicates whether a sentence represents the beginning of a new section of the text. The unitizing sentence code indicates whether a sentence is a summary or wrap-up sentence for a section of the text. Serial position codes the serial position of a sentence within each section of a text. Following Gernsbacher (1990), Stine-Morrow et al. (2008) argue that when readers begin a new section of a text, they are laying the foundation for a new mental representation of the text information. As the reader goes through that section, subsequent sentences become easier to integrate into the existing structure, so readers who create a coherent representation should speed up reading through the section. At the beginning of a new section, readers shift to a new mental structure, so reading slows down again.

Argument-Focused Processing Dimensions

We designed the support and refute dimensions to assess subjects' argument-focused processing. Individual subject regressions produce beta weights that indicate the extent to which each subject speeds up or slows down reading for each sentence as a function of the extent to which the sentence supports and refutes the topic proposition (e.g., TV violence causes real violence.) Support and refute beta weights are meaningful when crossed with text position or subject beliefs. With respect to text position, a positive support beta weight suggests *position consistent sentence processing* if the subject is reading a Yes text but *position inconsistent sentence processing* for a No text. A positive refute beta weight, in contrast, suggests position consistent sentence processing for a No text and position inconsistent processing for a Yes text. For beliefs, a positive support beta weight suggests *belief consistent sentence processing* if the subject is a believer and *belief inconsistent sentence processing* if the subject is a disbeliever. Finally, a positive refute beta weight suggests belief consistent sentence processing if the subject is a disbeliever and belief inconsistent sentence processing if the subject is a believer. It is also important to note that belief consistency at the sentence level is independent of the text position.

Support and refute ratings were collected in a separate experiment. One group of subjects provided support ratings and another provided refute ratings. Sentences appeared one at a time on the computer. For each sentence, subjects rated the extent to which the sentence supported or refuted the position. Supporting was defined as "to prove correct by evidence or argument." Ratings were provided on a 1–9 scale (1 = "Does not support claim," 5 = "Neither supports nor doesn't support claim," 9 = "Completely supports claim"). The rating scale remained on the screen throughout the task. Refuting was defined as "to prove wrong by evidence or argument." The rating scale for the refute subjects was 1 = "Does not refute claim," 5 = "Neither refutes nor doesn't refute claim," and 9 = "Completely refutes claim." Twelve subjects read and rated each sentence of each of the four texts. Each subject rated one evolution and one TV

violence text, with the text type and order counterbalanced across subjects. Data for one TV Yes and one TV No subject were lost. Support and refute ratings were calculated for each sentence of each text as the mean rating across subjects. The 10 most supporting and refuting sentences for the TV Yes and TV No texts are indicated in Appendix B.

Methods

Subjects. One hundred fifty subjects from a large midwestern U.S. university participated for credit in an Introductory Psychology course. Across eight conditions, seven had 13 subjects and one had 14 subjects.

Materials. Two texts were created for each topic, evolution and TV violence. The texts contain information found on the Internet and in journal articles, with additional writing and editing for cohesion. Each text addresses a proposition related to the topic. For evolution, both texts address the accuracy of the fossil record. The "Evolution Yes" text presents evidence that *all species evolved from a common ancestor due to inherited traits caused by natural selection*. The "Evolution No" text presents evidence that there are gaps in the fossil record and that it does not support evolutionary theory. For TV violence, both texts address longitudinal studies, studies in which data were collected before and after a town had access to TV and experimental studies (see Appendix B for TV violence texts.) The TV Yes text presents evidence suggesting that *watching TV violence causes real violence*. The TV No text presents evidence that most research on TV violence has flaws and that the preponderance of evidence does not support a causal link between TV and real violence. For all texts, the position is clearly stated in the title (e.g., "TV Violence Causes Real Violence") and in the first paragraph. Descriptive statistics for all texts are presented in Table 1.[2]

Design and procedure. Subjects reported beliefs and belief basis as part of an online prescreening survey within the first 2 weeks of the semester. For each

[2]A pilot study was conducted to ensure students understood that for each topic, the Yes texts support the positions and the No texts refute the positions. The texts were printed without titles. Thirty-six subjects read one evolution and one TV violence text, with the texts counterbalanced in terms of topic and order such that 18 subjects read and rated each text. Subjects were instructed to read the texts carefully for comprehension and that they would answer a question after reading. After reading, subjects responded to one question about the extent to which the text supported the proposition. Subjects responded on a 9-point scale (1 = "Completely refute (go against) the claim," 5 = "Neither support nor refute the claim," and 9 = "Completely support the claim"). The Evolution Yes text was rated as more supportive, $M = 7.06$, $SD = 1.98$, than the Evolution No text, $M = 3.00$, $SD = 1.88$, $F(1, 34) = 39.65$, $p < .0001$, $\eta_p^2 = .54$. The TV Yes text was also rated as more supportive, $M = 7.28$, $SD = 1.67$, than the TV No text, $M = 2.17$, $SD = 1.43$, $F(1, 34) = 97.35$, $p < .0001$, $\eta_p^2 = .74$.

TABLE 1
Text Characteristics and Mean Values (per Sentence) for Reading Time Predictors for the Experiment 1 Texts

	Evolution Yes	*Evolution No*	*TV Yes*	*TV No*
Text characteristics				
Words	1,883	1,872	1,783	1,899
Sentences	127	112	99	98
Paragraphs	29	36	25	26
Flesch-Kincaid grade level	10.2	11.0	13.1	11.0
Reading time predictor means				
Syllables	24.95	27.98	32.91	30.79
New arguments	1.59	1.56	1.55	1.35
Argument overlap	.32	.63	.88	.72
Topic sentence	.05	.04	.06	.06
Unitize sentence	.05	.03	.03	.03
Serial position	12.26	16.23	12.24	9.57
Support	4.91	3.79	5.46	4.46
Refute	4.11	5.57	3.87	5.11

topic, subjects indicated their belief in the proposition on a 9-point scale (1 = "completely disbelieve," 5 = "unsure whether I believe this," and 9 = "completely believe"). Subjects who responded 1–3 on the belief scale were classified as disbelievers and subjects who responded 7–9 were classified as believers. Belief basis was measured for each topic using Griffin's (2008) scale (Appendix A). Subjects with a positive score were considered evidence-based for that topic, and subjects with a negative score were considered affect-based for that topic. Subjects with a belief basis score of zero (N = 3 for evolution and N = 6 for TV violence) were not analyzed in belief basis calculations.

Each subject read one evolution and one TV violence text. Each text from one topic was paired with the texts from the other topic an equal number of times, and the order of the texts was counterbalanced across subjects. Subjects were run in groups of up to five and were seated at separate computer workstations. After providing informed consent, subjects read instructions stating they would read a text that presents evidence relating to one "side" of an issue. Subjects read the text one sentence at a time and were instructed to study each sentence until they understood it and then to press the spacebar to advance to the next sentence. Subjects then read the first assigned text. The text was displayed one paragraph at time with all letters converted to "-" and punctuation and spaces maintained. Sentences within a paragraph were displayed one at a time by the dashes changing into letters in a moving window fashion. After completing a paragraph, the first sentence of the next paragraph appeared along with the rest of the

paragraph as dashes. After reading the first text, subjects read the second text in the same fashion.

After the second text, subjects performed the cued-section recall task. Subjects were given the heading of the section and the approximate location and length of the section within the text (all of the cued sections were in the middle of their texts). Subjects were told to type as much as they could remember and to be as accurate and complete as possible. The recall task was untimed. The cued sections ranged from 14 to 21 sentences and from 226 to 315 words (see Appendix B). Upon completion of the first cued-section recall, subjects completed the second recall in the same manner. The entire experiment took approximately 40 minutes.

Results

We first describe the reading time analyses and then the cued-section recall analyses. For the reading time data, we began by analyzing total reading times within each topic as a function of subject beliefs and belief basis. There were no significant differences in total reading time as a function of beliefs or belief basis for either topic.

Sentence reading time analyses. Mean predictor values for the eight sentence-level predictors for each text are presented in Table 1. To analyze the sentence reading times, we followed the procedures outlined by Lorch and Myers (1990; Method 3). In the first step, we conducted a multiple regression for each subject and text. We regressed reading times for each sentence on the eight predictors simultaneously. The beta weights for each predictor in this analysis indicate the extent to which reading sped up or slowed down for each subject as a function of that predictor over and above the other seven predictors. These beta weights thus represent the unique reading time variance accounted for by each predictor for each subject. Table 2 shows the mean unstandardized beta weights for each predictor. In the second step, these beta weights were used as dependent variables to address two questions. First, a test of whether a beta weight for a predictor is different from zero across subjects represents a test of whether subjects tend to speed up or slow down as a function of that predictor over and above the other seven predictors (these results are also shown in Table 2). Second, group-level analyses were conducted in which differences in beta weights across levels of an independent variable represent tests of whether one group slows down or speeds up reading relative to another group.

For all analyses reported, we analyzed group differences in terms of subject beliefs. There were no significant differences between believers and disbelievers for either topic for any analysis, so beliefs per se are not discussed further. We also analyzed interactions between texts within each topic. In no case did the Yes

TABLE 2
Unstandardized Beta Weights, 95% Confidence Intervals, and Mean R^2 for Reading Time Predictors Regressed on Sentence Reading Times for Each Subject Separately as a Function of Text

	Evolution Yes ($n = 51$)		*Evolution No* ($n = 51$)		*TV Yes* ($n = 51$)		*TV No* ($n = 52$)	
Predictor	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*
Syllables	176.4[a]	[161.2, 191.7]	176.5[a]	[162.9, 190.0]	128.5[a]	[114.3, 142.8]	159.6[a]	[147.4, 171.8]
New arguments	211.4[a]	[109.5, 313.4]	46.9[b]	[1.5, 92.2]	119.5[c]	[41.4, 197.5]	207.4[a]	[137.1, 277.8]
Argument overlap	41.8	[−97.1, 180.7]	−251.2[a]	[−355.3, −147.2]	−125.4	[−352.4, 101.5]	−284.2[c]	[−455.3, −113.1]
Topic sentence	632.4[c]	[259.5, 1,005.4]	215.8	[−248.8, 680.4]	−545.4[b]	[−1,079.0, −11.9]	791.2[a]	[372.1, 1,210.3]
Unitize sentence	−351.0[c]	[−593.1, −109.0]	−1,017.9[a]	[−1,388.3, −647.5]	110.2	[−477.2, 697.6]	215.1	[−393.2, 823.4]
Serial position	−17.2[a]	[−26.2, −8.1]	−8.4	[−17.0, 0.1]	−34.8[a]	[−48.6, −21.0]	−3.5	[−16.5, 9.5]
Support	−20.2	[−117.5, 77.1]	300.1[a]	[122.4, 477.8]	222.5[a]	[143.4, 301.7]	−18.7	[−100.7, 63.2]
Refute	509.8[a]	[289.8, 729.8]	289.7[a]	[179.3, 400.2]	402.1[a]	[257.6, 546.5]	−203.1[b]	[−399.1, −7.2]
Mean R^2	.47		.54		.51		.53	

[a] $p < .001$. [b] $p < .05$. [c] $p < .01$.

or No text within each topic interact with variables of interest, so results are presented for the TV violence and evolution topics by combining the Yes and No texts within each topic.

Textbase and situation model dimensions. For the textbase and situation model dimensions, we address text-level belief consistency and belief basis effects. For the textbase variables, subjects reading the evolution texts slowed down more as a function of argument overlap if the text was inconsistent with their beliefs ($M = 278.4$, $SD = 395.3$) than if it was consistent with their beliefs ($M = 16.3$, $SD = 462.3$), $F(1, 82) = 9.76$, $p = .002$, $\eta_p^2 = .11$. There were no other belief consistency effects for either the evolution or TV violence texts. For the situation model variables, the evolution texts had one effect, which was that subjects reading a text that was inconsistent with their beliefs sped up more for unitizing sentences ($M = -1{,}093.4$, $SD = 1{,}282.1$) than subjects reading a text that was consistent with their beliefs ($M = -279.0$, $SD = 912.0$), $F(1, 82) = 11.41$, $p = .001$, $\eta_p^2 = .12$. There were no belief consistency differences for these situation model variables for the TV violence texts.

For belief basis, there were no textbase differences in sentence reading times as a function of belief basis. For the situation model measures, evidence-based evolution subjects reading the evolution texts sped up more as a function serial position ($M = -22.1$, $SD = 39.3$) than affect-based evolution subjects ($M = -3.5$, $SD = 18.1$), $F(1, 97) = 9.16$, $p = .003$, $\eta_p^2 = .09$. There were no other belief basis differences for the evolution or TV violence texts. Overall, the textbase and traditional situation model measures suggest that subjects did track these dimensions of comprehension, consistent with prior reading time studies (e.g., Graesser et al., 1980). In terms of the current research questions, it appears that textbase and situation model dimensions are generally not moderated by beliefs or belief basis.

Argument-focused sentence processing. We address three questions with the beta weights for support and refute ratings: (1) Do support and refute ratings account for unique variance in sentence reading times? (2) Do subjects alter reading times to a greater extent as a function of position consistent versus position inconsistent sentences? (3) Do argument-focused sentence processing factors vary as a function of beliefs or belief basis? To address our first question, the mean beta weights for support and refute predictors presented in Table 2 indicate that, in general, they did account for unique sentence reading time variance. Refute ratings significantly predicted reading times for all four texts, and support ratings predicted reading times for the Evolution No and TV Yes texts. These results suggest that the support and refute ratings are empirically separable; therefore, it is sensible to combine them with text position and subject beliefs, as described earlier.

Data for reading time changes as a function of position consistency are presented in Table 3. Subjects slowed down reading as a function of position consistent argument strength for the evolution texts but not the TV violence texts. Position inconsistent argument strength is highly predictive of reading slowdowns for both texts, however. A difference score was calculated for each subject by subtracting the position inconsistent beta weight from the position consistent beta weight. The negative difference scores indicate that for both texts, reading slowdowns were greater as a function of position inconsistent argument strength than position consistent argument strength. Reading time changes as a function of belief consistency are also presented in Table 3. For the evolution texts, readers slow down as a function of both belief consistent and inconsistent argument strength but more so for belief inconsistent sentences. Belief consistency does not predict reading times for the TV violence texts, however.

Belief and belief basis predictions of sentence reading times were analyzed with mixed analyses of variance in which belief consistency was a within-subjects variable and belief basis was a between-subjects variable. Data are shown in Figure 2. For evolution, there was a main effect of belief consistency in which beta weights for belief inconsistent sentence processing ($M = 359.3$) were greater than beta weights for belief consistent sentence processing ($M = 178.3$), $F(1,80) = 6.03$, $p = .02$, $\eta_p^2 = .07$. For TV violence, there was no main effect of belief consistency ($M = 79.8$ for belief inconsistent sentence processing and $M = 103.8$ for belief consistent sentence processing $F < 1$.) For belief basis, the evolution beta weights were greater for evidence-based evolution subjects ($M = 415.0$) than affect-based evolution subjects ($M = 136.3$), $F(1,80) = 6.36$, $p = .01$, $\eta_p^2 = .07$. The interaction was not significant. For the TV violence texts, there was no main effect of belief basis ($M = 124.1$ for evidence-based TV violence subjects and $M = 50.0$ for affect-based TV violence subjects, $F < 1$) and no interaction.

Belief basis effects were also analyzed with mixed analyses of variance in which position consistent and inconsistent sentence processing was a within-subjects variable and belief basis was a between-subjects variable. Data for evolution and TV violence are shown in Figure 3. For evolution, there was a main effect of position consistency in which beta weights were greater for position inconsistent sentence processing ($M = 412.3$) than for position consistent sentence processing ($M = 134.4$), $F(1,97) = 20.68$, $p < .0001$, $\eta_p^2 = .18$. Similarly for TV violence, the beta weights for position inconsistent sentence processing were greater ($M = 191.3$) than for position consistent sentence processing ($M = 11.7$), $F(1,95) = 14.97$, $p < .0001$, $\eta_p^2 = .14$. For belief basis for the evolution texts, evidence-based evolution subjects had marginally greater beta weights ($M = 370.6$) than affect-based evolution subjects ($M = 178.0$), $F(1,97) = 3.89$, $p = .05$, $\eta_p^2 = .04$. There also was a significant interaction,

TABLE 3
Unstandardized Beta Weights and 95% Confidence Intervals for Support and Refute Predictors Combined According to Text Position and Belief Consistency

	Text Position						*Subject Beliefs*					
	Consistent		*Inconsistent*		*Difference*		*Consistent*		*Inconsistent*		*Difference*	
	M	*95% CI*	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*
Evolution	134.7[a]	[56.2, 213.3]	405.0[b]	[264.5, 545.4]	−267.6[b]	[−388.5, −146.7]	169.9[a]	[45.9, 293.9]	351.5[b]	[208.4, 494.7]	−181.6[c]	[−330.4, −32.8]
TV Violence	7.6	[−105.0, 120.3]	189.6[b]	[98.5, 280.7]	−182.0[b]	[−270.4, −93.6]	93.7	[−34.4, 221.7]	79.0	[−60.5, 218.5]	14.7	[−109.1, 138.4]

Difference scores are consistent minus inconsistent for each subject.
[a] $p < .01$. [b] $p < .001$. [c] $p < .05$.

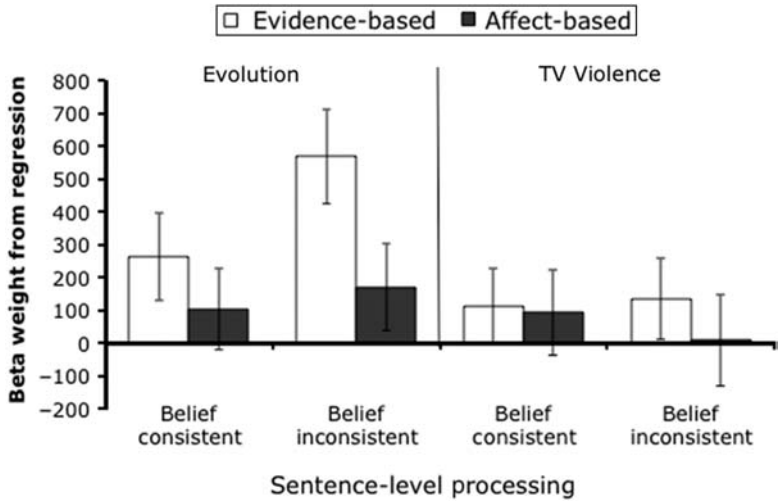

FIGURE 2 Beta weights from individual subject regressions showing changes in sentence-level reading times predicted by the extent to which sentences were consistent or inconsistent with subject beliefs. Sentence-level beta weights were created by matching support and refute beta weights for each subject with their belief status for each topic. Error bars are standard error of the means.

$F(1,97) = 6.12$, $p = .02$, $\eta_p^2 = .06$. As shown in Figure 3, the extent to which the position inconsistent beta weights are greater than the position consistent beta weights is larger for the evidence-based evolution subjects than for the affect-based evolution subjects. For the TV violence texts, there was no significant difference between evidence-based TV violence subjects ($M = 140.0$) and affect-based TV violence subjects ($M = 41.8$, $F < 1$). There also was no significant interaction between position consistency and belief basis ($F < 1$).

Cued section recall. To code the cued section recall data, the section of the text subjects were supposed to recall was first divided into clauses (Trabasso & Magliano, 1996). A clause is a standalone idea, typically consisting of a main verb and accompanying content. Recall protocols were also divided into clauses. Recall clauses were coded at the gist level. If a clause matched a clause from the cued section, it was coded as an instance of *cued section* recall. If not, the clause was compared with the rest of the text, and if it matched a clause it was labeled as an *other text* recall. If the recalled clause did not match any content from the text, it was labeled a *prior knowledge* clause. The prior knowledge code consisted primarily of general knowledge associations, with a small number of relevant prior knowledge connections. Agreement on the number of clauses for Evolution Yes, Evolution No, TV Yes, and TV No was 91%, 94%, 96%, and 80%,

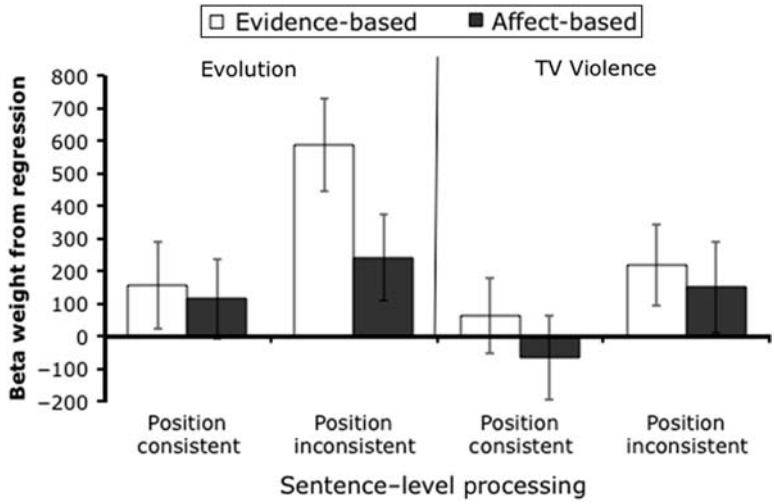

FIGURE 3 Beta weights from individual subject regressions showing changes in sentence-level reading times predicted by the extent to which sentences were consistent or inconsistent with the position argued for in the one-sided text. Sentence-level beta weights were created by matching support and refute beta weights for each subject with the text position for each topic. Error bars are standard error of the means.

respectively. Kappa was calculated from agreed upon clauses and was .76, .68, .71, and .84, respectively.

We were interested in recall performance as a function of subjects' belief consistency and belief basis. For all analyses, data were collapsed across the Yes and No text within each topic. For belief consistency, there were no significant differences for any of the three types of recall. Data by belief basis are presented in Table 4. Subjects reading the evolution texts had more recall of other text content if they were evidence-based evolution subjects than if they were affect-based evolution subjects, $F(1, 97) = 7.32$, $p = .008$, $\eta_p^2 = .07$. No other recall categories were significantly different as a function of belief basis.

Discussion

In Experiment 1 we found, consistent with our predictions, that subjects engage in argument-focused processing; a selective reading slowdown occurred as a function of the strength of arguments during comprehension of one-sided argumentative texts. These slowdowns were over and above the processing associated with traditional text processing factors. Subjects appear to be sensitive to the side of an argument that is taken in a particular sentence and how that

TABLE 4
Mean Cued-Section Recall and 95% Confidence Intervals as a Function of Topic and Belief Basis

	Evolution				*TV Violence*			
	Evidence-Based ($n = 49$)		*Affect-Based* ($n = 50$)		*Evidence-Based* ($n = 58$)		*Affect-Based* ($n = 37$)	
Recall Category	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*	*M*	*95% CI*
Cued section	2.08	[1.57, 2.59]	2.60	[2.08, 3.12]	.91	[0.55, 1.27]	.84	[.28, 1.39]
Other text	4.04[a]	[3.22, 4.86]	2.62	[1.96, 3.28]	5.07	[3.80, 6.34]	5.62	[4.62, 6.98]
Prior knowledge	2.27	[1.57, 2.96]	1.94	[1.48, 2.40]	2.45	[1.97, 2.93]	1.95	[1.36, 2.53]
Total	8.39	[7.14, 9.63]	7.16	[6.37, 7.95]	8.43	[7.28, 9.58]	8.41	[7.32, 9.49]

[a] $p < .01$.

sentence relates to the text position as a whole. This argument-focused processing was evident for both position consistent and position inconsistent sentences but was more pronounced for position inconsistent sentences. These results are consistent with the results of Wolfe, Britt, and Butler (2009) in suggesting that subjects are sensitive to other side information in argumentative texts. The current data expand on Wolfe et al.'s findings by indicating that this sensitivity is present during text processing and that in fact other side information may even be emphasized during processing. It is also important to note that by factoring out instances of argument overlap and the number of new argument nouns, it seems unlikely that subjects slowed down their reading merely because of breaks in the coherence of the text (Albrecht & O'Brien, 1993). The results are also independent of the extent to which the sentences are consistent with subjects' beliefs, consistent with the argument generation data of Wolfe and Britt (2008).

The argument-focused processing results appear to be consistent with the balanced evidence processing hypothesis introduced earlier. Subjects' position inconsistent processing suggests they attempt to understand both sides of contentious issues in an effort to create a relatively balanced situation model with respect to the evidence. An alternative explanation, however, is that the reading time results represent a metacognitive phenomenon. Subjects may believe that good argumentative texts only include position consistent information (Wolfe & Britt, 2008). If subjects are not expecting position inconsistent arguments, they may slow down reading because of confusion. By this confusion hypothesis, subjects would still form a situation model consistent with the one-sided macrostructure, and position inconsistent arguments would be fragmented in terms of their organization and poorly remembered.

The cued-recall data provide some ability to address the nature of the situation model formed by the subjects. If the balanced evidence processing hypothesis is correct, then parts of the text that represent arguments counter to the text macrostructure may be organized together. In the cued-recall task, this type of organization would make it more difficult for subjects to successfully access information from a particular section of the text without having intrusions from other parts of the text (Mannes & Kintsch, 1987). In the cued-recall task for the evolution texts, more clauses were recalled that were intrusions from other parts of the text than there were clauses from the to-be-recalled section. However, this result does not provide strong support for a balanced situation model, because the content is different in to-be-recalled sections and other text sections.

We were also interested in belief consistency and belief basis as potential moderators of argument-focused processing. For belief consistency, results were largely unsupportive of the possibility that subjects alter their processing based on whether they do or do not believe what they are reading. For the evolution texts, subjects slowed down reading more for sentences that were inconsistent with their beliefs, but this result did not hold for the TV violence texts. Subjects

also did not differ in any consistent way in terms of the textbase and traditional situation model measures, and there were no cued-recall differences as a function of belief consistency. These results are only partially supportive of the defense against attack hypothesis (Eagly et al., 2000). It is unclear whether our results contradict Maier and Richter's (2013) findings of better situation model memory for belief consistent information and better textbase memory for belief inconsistent information when reading multiple conflicting texts. Processing and memory may differ with multiple conflicting texts compared with the single texts used in the current experiment. In fact, Maier and Richter found that the presentation order of the texts influenced the belief consistency memory findings. This text order effect suggests that the manner in which arguments are presented is important. The texts in the current experiment, with position consistent and inconstant content integrated into the same text, could not be sensibly separated in the same manner.

For belief basis, we found some evidence that subjects process and recall these texts differently as a function of why they claim to hold their beliefs. For processing, there was an interaction for the evolution texts between belief basis and position consistency. Evidence-based evolution subjects displayed balanced evidence processing to a greater degree than affect-based evolution subjects. For cued-recall, evidence-based evolution subjects had more "other text" intrusions than affect-based evolution subjects for the evolution texts. If evidence-based subjects are making a greater effort than affect-based subjects to understand and integrate the position inconsistent arguments, then we would expect more recall intrusions from sections of the text other than the section that was assigned for recall. For the TV violence texts, there were no belief basis differences in processing or cued-recall. Accurate recall for the cued-recall section was extremely low overall, however. It is possible that the particular section subjects recalled was too difficult to access because of some characteristic of the texts themselves.

EXPERIMENT 2

In Experiment 2, we further examine the representation readers form after reading a one-sided scientific text about a disputed topic. We were particularly interested in the extent to which readers' situation models vary as a function of subjects' differences in argument-focused processing. For beliefs, if subjects engage in more belief inconsistent argument-focused processing, then we expect them to create situation models that are biased in favor of belief inconsistent evidence. For belief basis, if evidence-based subjects engage in more balanced evidence processing, then we expect them to create situation models that are more balanced than those created by affect-based subjects. In Experiment 2,

subjects read a one-sided text, then a neutral text, and then wrote a summary of the neutral text (this method is modeled after the text summary task of Mannes & Kintsch, 1987). In this task, the relative extent to which subjects' summaries of the neutral text reflects one-sided bias provides evidence about the nature of the representation of the one-sided text they read. Following Mannes and Kintsch, we reasoned that subjects would have a hard time reading two texts on the same topic and writing a summary about one without some interference from the content of the other one. As a result, the more one-sided the subjects' representation of the one-sided text, the more biased their summary of the neutral text would be. In addition, half the subjects wrote their summary in the same session as their comprehension and half wrote their summary after a 2-day delay.

We chose this task for a few reasons. First, we wanted a writing task that was neutral with respect to the topic. We reasoned that writing a summary of a one-sided text would be a poor method for assessing the representation of that text, because it is likely that the summary would only reflect position consistent information. Likewise, if subjects wrote an argumentative text, their argument schema may dictate that they only include position consistent information (Wolfe & Britt, 2008). With the neutral summary task, any one-sided bias in the summaries is not part of the instructions and therefore affords information about the representation of arguments from reading the first text. Second, this design allowed us to examine one-sided bias in subjects' representations as a function of belief consistency and belief basis. For example, we predict that evidence-based subjects will form more balanced situation models across position consistent and inconsistent evidence. Support for this hypothesis will be found if evidence-based subjects write less biased summaries compared with affect-based subjects. Third, we wanted subjects to engage in a relatively ecologically valid task, and a summary is a common task for students to write. We also increased ecological validity by having subjects read one paragraph at a time.

In this experiment, we used only the TV violence topic from Experiment 1. We did attempt to create a neutral text for the evolution topic but found that subjects consistently rated all of our neutral texts about fossil collection as pro-evolution. This situation may arise from the unique circumstances with the evolution topic, that even though the issue is disputed among the general public, the basics of evolution are not disputed in the scientific community. However, the TV violence topic provides a good test for our questions because the pattern of results was somewhat weaker than for evolution in Experiment 1, particularly for belief basis.

Methods

Subjects. One hundred forty-seven subjects from a large midwestern U.S. university participated for credit in an Introductory Psychology course. Five

subjects failed to show up for the 2-day delay and one was lost to computer error. The remaining 141 subjects were distributed across the eight conditions that cross text type, delay, and belief category, with 15 to 21 subjects per condition. The numbers of subjects in each condition were slightly different due to some differences in sign-ups as a function of subject beliefs.

Materials. The TV Yes and TV No texts from Experiment 1 were used. A third text, "TV Neutral," was created using the same procedure as the Experiment 1 texts. This text describes some of the issues involved with TV violence research and some of the studies described in the one-sided texts. The TV Neutral text does not provide results or conclusions for the experiments, however. The text contains 1,867 words, 107 sentences, 17.2 words per sentence, 20 paragraphs, and a Flesch-Kincaid grade level score of 12.7.

Design and procedure. Subjects' reported belief and belief basis for TV violence using the same prescreening process as Experiment 1. Believers and disbelievers were invited by e-mail to participate. After providing informed consent, subjects read the TV Yes or TV No text, which was crossed with subject beliefs. Instructions were the same as Experiment 1, with the exception that subjects read the text one paragraph at a time. When the subjects finished reading a paragraph, they pressed the spacebar to advance to the next paragraph. Upon completion of the one-sided text, subjects completed 10 unrelated puzzles that progressed automatically at 1-minute intervals. Next, all subjects read the TV Neutral text in the same fashion as the one-sided text. Then the subjects completed 10 more puzzles at 1 minute each. At this point, half of the subjects completed the summary task and half were dismissed and returned 2 days later to complete the summary task. Delay condition was crossed with text and belief such that a roughly equal number of subjects were in all cells of the 2 (text) × 2 (belief) × 2 (delay) design.

In the summary task, subjects wrote a summary of the TV Neutral text. Subjects were provided the title ("Television and Violence Research") as a cue and were instructed to write a summary of only that text. Subjects were instructed that the summary should describe the sections of the text in the order they appeared. Subjects were instructed that "A good summary should include: 1. A statement about the overall thesis or point of the text. 2. A description of each of the sections of the text and what points they were making. 3. Enough information that the reader can get a sense of what the text was about, but not so much that it seems like you are trying to remember every detail." Subjects were also explicitly told not to include information from the first text in the summary. Subjects typed their summaries into a Word file. The summary task was not timed, and the entire experiment took approximately 50 minutes.

Results

Our primary question to address is the extent to which summaries of the neutral text contain evidence of biased intrusions from the one-sided text that was read first. We analyzed intrusions by hand-coding summaries and in a separate rating experiment in which a new set of subjects rated the summaries for one-sided bias.

Hand coding of summaries. Summaries were first broken into clauses, consistent with the recall data in Experiment 1. Clauses were matched to one of five sources: *Neutral text* statements matched a specific clause from the neutral text. *Neutral macro* statements were gist or summary statements that were consistent with the neutral text but did not match a specific clause from the text. *One-sided text* statements matched a specific clause from the one-sided text, and *one-sided macro* statements were gist or summary statements that were consistent with the one-sided text but did not match a specific clause. Finally, if a clause did not match any of the above categories, it was coded as *prior knowledge*. As with Experiment 1, almost all prior knowledge clauses represented general world knowledge, not specific content knowledge. Coding was done by two raters for the first 10 subjects of each text ($\kappa = .81$ for TV No and .77 for TV Yes.)

In addition to these five categories, we also created a measure designed to better capture the overall level of one-sided bias in the summaries. The one-sided macro statements are most commonly statements that represent global bias in the summaries. For example, one student wrote, "The text 'Television and Violence Research' main purpose was to show that studies have found a correlation between television and violence." This clause does not match a clause from either text but is clearly consistent with the one-sided (Yes) text and not the neutral text. To calculate the relative amount of bias in each summary, we first computed a normalized score for both one-sided macro statements and for the sum of the neutral text and neutral macro statements. The difference between these two normalized scores, which we call *relative bias*, reflects the relative extent to which summaries have clearly biased compared with neutral statements. We used normalized scores because neutral text statements were more frequent than one-sided macro statements.

Table 5 shows the mean number of statements in each coding category as a function of delay. Subjects wrote more neutral text statements after the 10-minute delay compared with the 2-day delay, $F(1, 137) = 19.93$, $p < .0001$. After the 2-day delay, subjects wrote more one-sided text statements, $F(1, 137) = 7.45$, $p = .007$, more one-sided macro statements, $F(1, 137) = 11.22$, $p = .001$, and more prior knowledge statements, $F(1, 137) = 4.54$, $p = .03$. Subjects also had higher scores on the relative bias measure after the 2-day delay than the 10-minute delay, $F(1, 137) = 24.60$, $p < .0001$.

TABLE 5
Clause Category Means, 95% Confidence Intervals, and Effect Sizes From Summary Statements as a Function of Delay Condition

	Ten-Minute		*Two-Day*		
Clause Category	*M*	*95% CI*	*M*	*95% CI*	η^2
Neutral text	4.73[a]	[4.15, 5.30]	2.83	[2.22, 3.44]	.13
Neutral macro	1.41	[1.13, 1.67]	1.12	[0.83, 1.41]	.01
One-sided text	2.38	[1.85, 2.93]	3.47[b]	[2.90, 4.04]	.05
One-sided macro	.58	[0.31, 0.84]	1.23[b]	[0.95, 1.51]	.08
Prior knowledge	2.97	[2.05, 3.89]	4.40[c]	[3.44, 5.37]	.03
Relative bias	−.58	[−0.90, −0.25]	.61[a]	[0.26, 0.95]	.15

[a] $p < .001$. [b] $p < .01$. [c] $p < .05$.

In terms of beliefs themselves, believers wrote more prior knowledge statements ($M = 4.62$, $SD = 5.04$) than disbelievers ($M = 2.78$, $SD = 2.78$), $F(1, 137) = 7.46$, $p = .007$, $\eta_p^2 = .05$. No other variables differed as a function of beliefs, and beliefs did not interact with delay for any variables. To address our primary questions, we examined summary statements from each of our coding categories as a function of belief consistency and belief basis. In each case, we examined main effects and interactions with delay condition. For belief consistency, there were no main effects or interactions with delay. For belief basis, evidence-based TV violence subjects wrote summaries with marginally lower scores on the relative bias measure ($M = -0.19$, $SD = 1.37$) than affect-based TV violence subjects ($M = 0.35$, $SD = 1.98$), $F(1, 131) = 3.25$, $p = .07$, $\eta_p^2 = .02$. The only other effect of belief basis was an interaction with delay for the number of prior knowledge statements, $F(1, 131) = 5.94$, $p = .02$, $\eta_p^2 = .04$. Evidence-based TV violence subjects wrote more prior knowledge statements after the 2-day ($M = 4.91$, $SD = 4.17$) compared with the 10-minute delay ($M = 2.24$, $SD = 1.75$). Affect-based TV violence subjects wrote fewer prior knowledge statements after the 2-day ($M = 3.60$, $SD = 3.20$) compared with the 10-minute delay ($M = 4.57$, $SD = 7.09$). We also conducted correlational measures to specifically address belief basis effects. For analysis purposes, we treated belief basis as a categorical variable. But because the belief basis scale is continuous, we correlated belief basis with the five coding category variables and with the relative bias measure. Belief basis was not correlated with any of the five coding category variables but was negatively correlated with the relative bias measure, $r(141) = -.18$, $p = .03$. Thus, as subjects were more evidence-based with respect to TV violence, they wrote less biased summaries.

Subjects' ratings of summaries. Our goal was to assess the overall level of one-sided bias in the summaries, so a separate group of subjects rated the summaries for bias. Instructions stated that the subjects would read summaries of

articles that were written by students as part of a separate experiment. Rating subjects were told that in the original experiment subjects read more than one article on TV research and wrote a summary of one of the articles. Rating subjects were also instructed that the original subjects read articles that supported the proposition, opposed the proposition, or were neutral. For each summary, rating subjects made two judgments. The first judgment was whether the *article* the student in the original experiment summarized presented evidence that supported or refuted the proposition or whether the article was neutral. Subjects made a rating on a 9-point scale (1 = "text refuted the claim," 5 = "text was neutral about the claim," and 9 = "text supported the claim"). The second judgment was whether the *person* writing the summary believes that TV violence causes real violence. This rating was also made on a 9-point scale (1 = "person disbelieves the claim," 5 = "person is neutral about the claim," and 9 = "person believes the claim"). Subjects were further instructed that these ratings could be different because, for example, a student could write a summary of a one-sided article without personally believing the content of the article. We did not have specific predictions about differences between article and person ratings; we collected both ratings because we wanted to be sure to capture whatever one-sided bias may be in the summaries. Each rating subject read and rated 10 or 11 summaries. For counterbalancing, 10 different random orderings of all 141 subjects were produced. One rating subject read and rated the first 10 of the first random ordering, the next subject rated the next 10, and so on. A total of 140 subjects participated in the rating experiment, and each summary was rated by 10 subjects.

For all analyses, we subtracted 5 from the bias ratings so that a value of 0 represents a neutral summary, a positive value represents a "Yes bias" and a negative value represents a "No bias." The first text that was read (Yes or No) had large overall effects on summaries. Subjects who read the TV Yes text first wrote summaries that were higher in article bias ($M = 1.55$, $SD = 1.12$) and person bias ($M = 0.73$, $SD = 1.31$) than subjects who read the TV No text first (article bias $M = -0.67$, $SD = 1.22$; $F(1, 139) = 126.52$, $p < .0001$, $\eta_p^2 = .48$; person bias $M = -1.15$, $SD = 1.08$; $F(1, 139) = 86.51$, $p < .0001$, $\eta_p^2 = .38$). We did not predict an overall delay effect, but summaries written after a 2-day delay were more biased toward the proposition for both article ratings ($M = 0.12$, $SD = 1.46$ for 10-minute and $M = 0.87$, $SD = 1.69$ for 2-day; $F(1, 139) = 7.90$, $p = .006$, $\eta_p^2 = .05$) and for person ratings ($M = -0.48$, $SD = 1.35$ for 10-minute and $M = 0.16$, $SD = 1.64$ for 2-day; $F(1, 139) = 6.52$, $p = .01$, $\eta_p^2 = .05$).

To address our questions about belief consistency and belief basis, we conducted analyses of variance in which text effects were also included. For the summary ratings, believers did not differ significantly from disbelievers in terms of article bias, $F(1, 137) = 2.80$, ns, or in terms of person bias, $F(1, 137) = 2.15$, ns, and there was no interaction with text condition for either article or person bias (F's < 1.5). There also were no interaction effects between subject beliefs

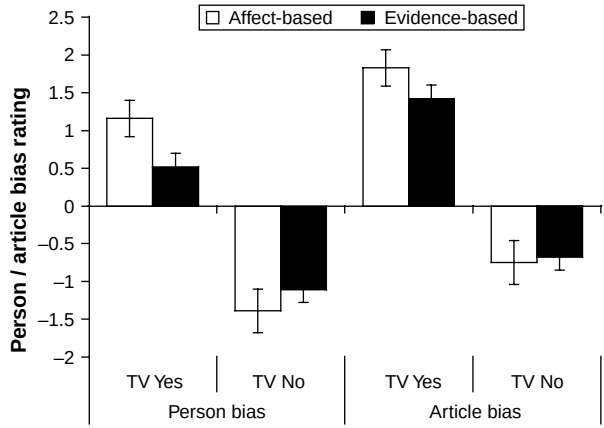

FIGURE 4 Person and article bias ratings of subjects' summaries of the TV Neutral text as a function of the first text read (TV Yes or TV No). Data are presented as a function of subjects' belief basis. Error bars are standard error of the means.

and delay condition, so there is no evidence that subjects' beliefs influence the extent to which they are successful at writing neutral summaries.

The potential influence of belief basis would manifest itself as an interaction between belief basis and text condition. If evidence-based TV violence subjects wrote less biased summaries, then they should have lower bias ratings when reading the TV Yes text first and higher ratings when reading the TV No text first (both closer to 0). Figure 4 shows the belief basis data. There were no main effects of belief basis on article bias or person bias (both F's < 1). The predicted interaction was significant for person bias ratings, $F(1, 131) = 4.28$, $p = .04$, $\eta_p^2 = .03$, but not for article bias ratings, $F(1, 131) = 1.20$, ns. Thus, for the ratings of the extent to which the person in the original experiment displayed bias in writing his or her summary, evidence-based TV violence subjects wrote more neutral summaries than affect-based TV violence subjects. In terms of the delay condition, we were interested in whether the belief basis $\times$ text interaction was different as a function of the delay. There were no significant three-way interactions between belief basis, text, and delay for either article bias or person bias.

Discussion

We assessed subjects' ability to study and summarize a neutral text about TV violence research after studying a one-sided text about the same topic. As expected, summaries showed evidence of bias from the one-sided text,

suggesting that subjects had a difficult time separating the neutral text content from the one-sided content. The *number* of intrusions from the one-sided text did not differ as a function of either beliefs or belief basis, however. This result suggests that the extent of content integration across texts was comparable as a function of our individual difference measures. If the number of intrusions is comparable across individual difference measures, then differences in one-sided bias in the summaries cannot be explained by the amount of integration of the neutral and one-sided content. As a result, bias as a function of beliefs or belief basis is interpreted as an indication of one-sided bias in the situation model that subjects created when reading the one-sided text.

One-sided bias in the summaries did not differ as a function of subjects' beliefs on any bias measure. Belief basis, however, was negatively correlated with the relative bias score, indicating that more evidence-based subjects wrote less biased summaries. Subject ratings of the summaries also indicated that evidence-based subjects wrote less biased summaries (as assessed by the person bias measure) than affect-based subjects, regardless of the text they read or their beliefs. Summaries written by affect-based subjects were more biased in the direction of the position of the one-sided text, regardless of whether that position was consistent with their prior beliefs. These results did not interact with delay. In the General Discussion, we discuss why the results are consistent with the balanced evidence processing hypothesis for the evidence-based subjects as explained in Experiment 1. The results are not consistent with the position consistent processing hypothesis.

GENERAL DISCUSSION

We examined students' processing and memory representation of the arguments presented in one-sided scientific texts about disputed topics. The texts were extended, naturalistic texts that mentioned and refuted evidence and arguments that were inconsistent with the main position in the text, as is characteristic of high quality argumentative texts (Wolfe, Britt, & Butler, 2009). Over and above textbase and situation model factors, subjects in Experiment 1 altered their processing as a function of the arguments in the texts. Support and refute ratings of the sentences provided unique predictions of sentence reading time variance. These results, which we refer to as *argument-focused processing*, suggest that subjects are sensitive to the match between the position being argued for in the text and the extent to which particular sentences support or refute that position. We elaborate on argument-focused processing and the situation model representations that result from this processing. We also discuss how argument-focused processing appears to be moderated by subjects' belief basis but not their beliefs.

Argument-Focused Processing

The reading time data from Experiment 1 suggest that subjects are capable of simultaneously tracking both position consistent and position inconsistent arguments as they read one-sided argumentative texts. We considered two possible processing hypotheses, *position consistent processing* and *balanced evidence processing*. The two hypotheses differ in that position consistent processing suggests that subjects will put processing effort into understanding the side of the text that is being argued for. This hypothesis arises from research suggesting that many subjects expect one-sided argumentative texts to contain only arguments that support the position (Kuhn et al., 1997; Wolfe & Britt, 2008; Wolfe, Britt, Petrovic, et al., 2009). The balanced evidence processing hypothesis suggests that subjects are engaged in an effort to understand and integrate information that is inconsistent with the text position. Overall, the reading slowdowns were greater as a function of position inconsistent argument strength compared with position consistent argument strength. These reading time data are more consistent with the balanced evidence processing hypothesis. The number of intrusions from other parts of the text on the cued-section recall task also suggest that subjects are attempting to integrate information across sections of the text, consistent with the balanced evidence processing hypothesis.

The reading time and memory data in the current study appear to be in conflict with research supporting the myside bias in argumentation (Kuhn et al., 1997; Kuhn & Udell, 2003; Wolfe & Britt, 2008). However, we believe the current data illustrate a difference between processing effort during argument comprehension and the use of evidence and reasons in argument generation tasks. In Wolfe and Britt's study, subjects chose to study both sides of a controversial issue, even though they generated arguments that primarily included only myside information. Other data on argument comprehension and use suggest that many students have a difficult time identifying opposing arguments in argumentative texts (National Center for Education Statistics, 2010; Wolfe, Britt, Petrovic, et al., 2009). Our data do suggest an ability on the part of subjects to differentiate position consistent and position inconsistent arguments. However, determining that a particular sentence is somehow related to an opposing argument is a simpler task than the formal process of identifying opposing arguments. For example, other research on argumentation indicates that students are relatively poor at determining whether reasons support claims in simple arguments (Britt et al., 2008; Larson, Britt, & Larson, 2004). In the current task, however, it is not necessary to formally distinguish the different parts of an argument, only to recognize whether sentences are generally related to the position consistent or inconsistent side. The argument quality ratings of Wolfe, Britt, and Butler (2009) are similar in that they suggest subjects' sensitivity to counterarguments but do not require subjects to specify the formal aspects of the arguments. Thus, part of

the novel contribution of this work is to add to the argumentation literature by suggesting that subjects attempt to comprehend both sides of an argument, even if that effort is not evident in argument evaluation or generation tasks.

Individual Differences in Argument Processing and Representation: Belief Basis

The other primary goals of this research were to examine the extent to which argument-focused processing, and subsequent representations, varied as a function of two individual difference variables. Reading time data in Experiment 1 and assessments of subjects' situation models in Experiments 1 and 2 suggest that the balanced evidence processing discussed in the previous section varies as a function of belief basis. Subjects who claimed to hold their belief for evidence-based reasons displayed more balanced evidence processing than subjects who claimed to hold their beliefs for affect-based reasons. Evidence-based evolution subjects in Experiment 1, compared with affect-based evolution subjects, had greater sentence reading time slowdowns for position inconsistent sentences for the evolution texts. For the cued-section recall task, evidence-based evolution subjects had more recall intrusions from sections of the text they were not supposed to recall (for the evolution texts; for the TV violence texts, recall of the correct section was poor overall). In Experiment 2, subjects read a one-sided text about TV violence and then read and summarized a neutral text. Evidence-based TV violence subjects wrote summaries that were less biased, whereas affect-based TV violence subjects wrote summaries that were biased in the direction of the text position. These results provide preliminary evidence that the processing engaged in by evidence-based subjects results in a situation model that is more balanced across arguments than a one-sided textbase representation that mirrors the macrostructure of the text. Affect-based subjects appear to create situation model representations that are more reflective of the one-sided text macrostructure. It is important to note, however, that for affect-based subjects to have a situation model that is consistent with the text macrostructure does not mean they have not put effort into processing the text. The reading time and memory results in Experiment 1 suggest that overall, affect-based subjects put effort into the traditional tasks that a successful reader must engage in: linking concepts across sentences, building an overall mental structure to represent the text organization, and so on (van Dijk & Kintsch, 1983).

At least two issues about balanced evidence processing and belief basis are important for future research. One issue is the extent to which balanced evidence processing represents a deliberate processing decision or if evidence- and affect-based subjects differ on a more fundamental aspect of reading, such as comprehension skill. We have some ability to assess this possibility with our Experiment 1 data by tabulating the number of subjects who were consistent in

their belief basis across topics compared with subjects who reported different belief bases for the topics. A total of 94 subjects were believers or disbelievers for both topics and had non-zero belief basis scores. Of those, 63 reported a consistent belief basis across the two topics, and 31 reported different belief bases. If belief basis differences are based on general reasoning or comprehension skill, we might expect fewer subjects to be inconsistent with their belief basis across topics. Griffin (2004) also found that across eight topics, there was wide variability within subjects in terms of belief basis. Thus, the current evidence suggests that belief basis is not a stable trait across topics but rather a factor that varies across topics within a person.

Another question for future research is whether prior knowledge may account for belief basis effects such as those reported here. We did find in Experiment 2 that subjects included more prior knowledge statements in their summaries after the 2-day delay than immediately. The vast majority of these statements were general knowledge statements, however, not topic-specific statements. It is possible, for example, that evidence-based subjects have more topic-specific knowledge than affect-based subjects. However, we believe it is unlikely that prior knowledge differences could explain the present results, because the current pattern of results do not appear to match typical knowledge effects. High knowledge subjects (compared with low knowledge subjects) typically recall more accurate text content and have fewer recall intrusions (Recht & Leslie, 1988; Spilich, Vesonder, Chiesi, & Voss, 1979), which we did not find as a function of belief basis. Wiley (2005) found differences in memory for information on opposing sides of an argument as a function of prior knowledge, but we did not find comparable belief basis effects in recall in Experiment 1. Under typical circumstances, high knowledge subjects also read faster than low knowledge subjects (Caillies, Denhiére, & Jhean-Larose, 1999), and we did not find overall reading speed differences as a function of belief basis. Finally, in related work dealing with TV violence and spanking as topics, Wolfe, Kurby, and Taylor (2012) found no relationship between prior knowledge and belief basis across three studies.

Belief Consistency

The general finding of a lack of belief consistency effects is consistent with the findings in the attitude and memory literature that often show little or no memory advantage for information that subjects believe (Eagly et al., 2000; Greenwald & Sakumura, 1967). However, in Experiment 1 for the evolution texts, one belief consistency effect did emerge at the argument processing level: Subjects slowed down reading more for sentences that were higher in belief inconsistency. This finding is consistent with the defense against attack hypothesis of Eagly et al. (2000) and with the attitude inconsistent reading time data of Edwards and Smith

(1996). No other results at the sentence processing level, or with respect to belief consistency of the text, however, indicated that subjects put more processing resources into belief inconsistent content. In fact, affect-based subjects reading the TV violence texts slowed down more as a function of both support and refute ratings if the text was consistent with their beliefs. Across the traditional textbase and situation model measures of processing, subjects did not differ as a function of beliefs. The measure of argument-focused processing that we developed and used in this research could be useful in future research on potential belief-related effects on processing. However, our current data do not suggest specific hypotheses related to beliefs.

Conclusions

We view the current research as an important step toward a broader understanding of how students comprehend science content. From a methodological standpoint, our use of sentence reading time data to assess argument-focused processing is one that could be applied to a number of different comprehension-related questions. In addition, the type of processing revealed by the evidence-based subjects is similar to elaborative processing activities that have been shown in previous research to be associated with a greater ability to transfer knowledge to new situations, and use it in a flexible manner to solve problems (Coté, Goldman, & Saul, 1998; Mannes & Kintsch, 1987; McNamara, 2004). Thus, from an educational standpoint, we should encourage and facilitate the processing activities of the evidence-based subjects more so than the affect-based subjects. It is a question for future research whether argument focused processing activities or belief basis can be manipulated through instructional programs.

ACKNOWLEDGMENTS

We thank Jill Hinton for helpful comments on an earlier version of the manuscript.

REFERENCES

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061–1070.

Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, *55*, 390–404.

Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2008). I agreed with what? Memory for simple argument claims. *Discourse Processes*, *45*, 52–84.

Britton, B. K. (1994). Understanding expository text: Building mental structures to induce insights. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). San Diego, CA: Academic Press.

Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist*, *56*, 477–489.

Caillies, S., Denhière, G., & Jhean-Larose, S. (1999). The intermediate effect: Interaction between prior knowledge and text structure. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 151–168). Mahwah, NJ: Lawrence Erlbaum Associates.

Centerwall, B. S. (1993). Television and violent crime. *Public Interest*, *111*, 56–71.

Coté, N. C., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1–53.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, *33*, 109–128.

Eagly, A. H., Chen, S., Chaiken, S., & Shaw-Barnes, K. (1999). The impact of attitudes on memory: An affair to remember. *Psychological Bulletin*, *125*, 64–89.

Eagly, A. H., Kulesa, P., Brannon, L. A., Shaw, K., & Hutson-Comeaux, S. (2000). Why counterattitudinal messages are as memorable as proattitudinal messages: The importance of active defense against attack. *Personality and Social Psychology Bulletin*, *26*, 1392–1408.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*, 5–24.

Freedman, J. L. (2007). Television violence and aggression: Setting the record straight. Retrieved from http://www.mediainstitute.org/IssuePapers.php

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, *19*, 135–151.

Greenwald, A. G., & Sakumura, J. S. (1967). Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology*, *7*(4, Pt. 1), 387–397.

Griffin, T. D. (2004). *Revising beliefs based in evidence versus affect: Effects on knowledge acquisition and conceptual change*. Unpublished doctoral dissertation, University of Illinois, Chicago.

Griffin, T. D. (2008). Faith: Serving emotional epistemic goals rather than evidence coherence. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 2059–2064). Austin, TX: Cognitive Science Society.

Griffin, T. D., & Ohlsson, S. (2001). Beliefs versus knowledge: A necessary distinction for explaining, predicting, and assessing conceptual change. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third conference of the Cognitive Science Society* (pp. 392–397). Mahwah, NJ: Lawrence Erlbaum Associates.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, *15*, 287–315.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*, 1245–1260.

Larson, M., Britt, M. A., & Larson, A. A. (2004). Disfluencies in comprehending argumentative texts. *Reading Psychology*, *25*, 205–224.

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measured data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 149–157.

Maier, J., & Richter, T. (2013). Text-belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, *31*, 151–175.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction*, *4*, 91–115.
McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.
Murphy, P. K., & Mason, L. (2006). Changing knowledge and beliefs. In P. A. Alexander & P. H. Winne (Eds.), *The handbook of educational psychology* (pp. 305–324). Mahwah, NJ: Lawrence Erlbaum Associates.
National Center for Education Statistics. (2010). *The nation's report card: Grade 12 reading and mathematics 2009 national and pilot state results* (NCES 2011-455). Washington, DC: U.S. Department of Education.
Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory for text. *Journal of Educational Psychology*, *80*, 16–20.
Roberts, J. V. (1985). The attitude-memory relationship after 40 years: A meta-analysis of the literature. *Basic & Applied Social Psychology*, *6*, 221–241.
Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 275–290.
Stine-Morrow, E. A. L., Miller, L. M. S., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging*, *23*, 131–153.
Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, *21*, 255–287.
van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.
Wiley, J. (2005). A fair and balanced look at the news: What affects memory for controversial arguments? *Journal of Memory and Language*, *53*, 95–109.
Wolfe, C. R., & Britt, M. A. (2008). The locus of the myside bias in written argumentation. *Thinking & Reasoning*, *14*, 1–27.
Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, *26*, 183–209.
Wolfe, C. R., Britt, M. A., Petrovic, M., Albrecht, M., & Kopp, K. (2009). The efficacy of a web-based counterargument tutor. *Behavior Research Methods*, *41*, 691–698.
Wolfe, M. B. W., Kurby, C. A., & Taylor, A. R. (2012, July). *Argument evaluation and recall as a function of belief in the argument*. Poster presented at the twenty-second annual conference of the Society for Text & Discourse, Montreal, Canada.
Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situations model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386–397.
Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, *2*, 199–220.

APPENDIX A: GRIFFIN'S (2008) BELIEF BASIS SCALE (APPLIED TO TV VIOLENCE)

(A) When it comes to issues like television violence, I use my heart and not my head to tell me the truth.
(E) My opinion about television violence is based on carefully examining the evidence both for and against it.
(A) I don't need "proof"; I have faith that my opinion about television violence is correct.

(A) My opinion about television violence makes me feel good or is comforting.
(E) My opinion about television violence is supported by scientific evidence.

Subjects responded to each question on a 9-point scale (1="Not at all my reason," 5 = "Unsure whether this is my reason," and 9 = "Completely my reason"). The belief basis score for each subject with respect to each topic is the difference between the mean of the two evidence questions (E) and the mean of the three affect questions (A). Subjects with a positive score were considered evidence-based, and subjects with a negative score were considered affect-based.

APPENDIX B: TV YES AND TV NO TEXTS

Underlined sentences indicate the 10 sentences in each text that are most supportive of the proposition "watching television violence causes real violence." Italicized sentences indicate the 10 sentences that most strongly refute the proposition. Parts of the TV Yes text come from Bushman and Anderson (2001) and Centerwall (1993). Parts of the TV No text come from Freedman (2007). For the cued-section recall task, subjects recalled the section "Longitudinal Study" for both texts.

TV Yes Text

Television Violence Causes Real Violence

Psychologists have studied the effect of violent media on aggression for several decades. Hundreds of studies have been conducted on this topic. Scientific evidence from a collection of studies, such as those on media-related aggression, can be integrated and summarized in a narrative review or in a meta-analytic review. Both types of review have been conducted on the research literature about media violence and aggression. All have come to the same conclusion, which is that viewing violence increases aggression.

The Seeds of Aggression

Children are born ready to imitate adult behavior. That they can, and do, imitate an array of adult facial expressions has been demonstrated in newborns as young as a few hours old. This is before they are even old enough to know that they have facial features. It is a most useful instinct, for the developing child must learn and master a vast repertoire of behavior in short order.

Children have an instinctive desire to imitate. But they do not possess an instinct for determining whether a behavior ought to be imitated. They will imitate anything, including behavior that most adults regard as destructive and antisocial. It may give pause for thought, then, to learn that infants as young as fourteen months demonstrably observe and incorporate behavior as seen on television.

Naturally, as children get older, they come to know better. But their earliest and deepest impressions are laid down at an age when they still see television as a factual source of information about the outside world. In that world, violence is common. In addition, the commission of violence is generally powerful, exciting, charismatic, and

effective. In later life, serious violence is most likely to erupt at moments of severe stress. It is precisely at such moments that adolescents and adults are most likely to revert to their earliest, most visceral sense of the role of violence in society and in personal behavior. Much of this sense will have come from television.

With and Without Television

In 1973, a remote rural community in Canada acquired television for the first time. The acquisition of television at such a late date was due to problems with signal reception rather than any hostility toward TV. As reported in The Impact of Television (1986), Tannis Williams and her associates at the University of British Columbia investigated the effect of television on the children of this community (which they called "Notel"). For comparison, they also looked at two similar towns that already had television.

The researchers observed forty-five first- and second-graders in the three towns for rates of inappropriate physical aggression before television was introduced into Notel. Two years later, the same forty-five children were observed again. To prevent bias in the data, the research assistants who collected the data were kept uninformed as to why the children's rates of aggression were of interest. Furthermore, a new group of research assistants was employed the second time around. This ensured that the data gatherers would not be biased by recollection of the children's behavior two years earlier.

Rates of aggression did not change in the two control communities. By contrast, the rate of aggression among Notel children increased by 160 percent. The increase was observed in both boys and girls, in those who were aggressive to begin with and in those who were not. Television's enhancement of aggression was entirely general and not limited to a few "bad apples."

In another Canadian study, Gary Granzberg and his associates at the University of Winnipeg investigated the impact of television upon Indian communities in northern Manitoba. As described in Television and the Canadian Indian (1980), forty-nine third-, fourth-, and fifth-grade boys living in two communities were observed from 1973, when one town acquired television, until 1977, when the second town did as well. The aggressiveness of boys in the first community increased after the introduction of television. The aggressiveness of the boys in the second community, which did not receive television then, remained the same. Television was later introduced in the second community. The observed levels of aggressiveness then increased there as well.

White South Africans have lived in a prosperous, industrialized society for decades. But they did not get television until 1975 because of tension between the Afrikaner- and English-speaking communities. The country's Afrikaner leaders knew that a South African television industry would have to rely on British and American shows to fill out its programming schedule. They felt that this would provide an unacceptable cultural advantage to English-speaking South Africans. So, rather than negotiate a complicated compromise, the government simply forbade television broadcasting. The entire population of two million whites was thus excluded from exposure to television for a quarter century after the medium was introduced in the United States.

In order to determine whether exposure to television is a cause of violence, homicide rates were compared in South Africa, Canada, and the United States. *Blacks in South Africa live under quite different conditions than blacks in the United States.* Because of this, the comparison was limited to white homicide rates in South Africa and the United States, and the total homicide rate in Canada (which was 97 percent white in 1951). The

homicide rate was chosen as a measure of violence because homicide statistics are exceptionally accurate.

From 1945 to 1974, the white homicide rate in the United States increased 93 percent. In Canada, the homicide rate increased 92 percent. In South Africa, where television was banned, the white homicide rate declined by 7 percent.

Longitudinal Study

In another study conducted from 1960 to 1981, Eron and Huesmann followed 875 children living in a semi rural U.S. county. Eron and Huesmann found that for both boys and girls, the amount of television watched at age eight predicted the seriousness of criminal acts by age thirty. This remained true even after controlling for the children's baseline aggressiveness, intelligence, and socioeconomic status. Eron and Huesmann also observed second-generation effects. Children who watched more television at age eight later, as parents, punished their own children more severely than did parents who watched less television as children. Second- and now third-generation effects are accumulating at a time of unprecedented youth violence.

All of the U.S. and Canadian studies of prolonged childhood exposure to television demonstrate a positive relationship between exposure and physical aggression. The critical period is preadolescent childhood. Later exposure does not appear to produce any additional effect. However, the aggression enhancing effect of exposure in pre-adolescence extends into adolescence and adulthood. This suggests that any interventions should be designed for the children and their caregivers rather than for the general adult population.

These studies confirmed the beliefs of most Americans. According to a Harris poll at the time of the studies, 43 percent of American adults believe that television violence "plays a part in making America a violent society." An additional 37 percent think it might.

The Television Industry Takes a Look

CBS commissioned William Belson to undertake what would be the largest and most sophisticated study yet, an investigation involving 1,565 teenage boys. In Television Violence and the Adolescent Boy (1978), Belson controlled for one hundred variables. He found that teenage boys who had watched above-average quantities of television violence before adolescence were committing acts of serious violence (e.g., assault, rape, major vandalism, and abuse of animals) at a rate 49 percent higher than teenage boys who had watched below-average quantities of television violence. CBS had invested a large sum of money. However, the executives were notably unenthusiastic about the report.

ABC commissioned Melvin Heller and Samuel Polsky of Temple University to study young male felons imprisoned for violent crimes (e.g., homicide, rape, and assault). In two surveys, 22 and 34 percent of the young felons reported having consciously imitated crime techniques learned from television programs. The more violent of these felons were the most likely to report having learned techniques from television. Overall, the felons reported that as children they had watched an average of six hours of television per day. Six hours is approximately twice as much as children in the general population at that time.

NBC relied on a team of four researchers, three of whom were employees of NBC. Indeed, the principal investigator, Ronald Milavsky, was an NBC vice president. The team observed 2,400 schoolchildren for up to three years to see if watching television violence increased their levels of physical aggressiveness. *In Television and Aggression (1982),*

Milavsky and his associates reported that television violence had no effect upon children's behavior. However, independent investigators have examined their data. They all concluded that, to the contrary, the data show that television violence did cause a modest increase of about 5 percent in average levels of physical aggressiveness. When pressed on the point, Milavsky and his associates conceded that their findings were consistent with the conclusion that television violence increased physical aggressiveness "to a small extent." *They did not concede that television violence actually caused an increase. They only conceded that their findings were consistent with such a conclusion.*

The NBC study results raise an important objection to the conclusion that TV violence causes aggression. Studies have repeatedly demonstrated that childhood exposure to television increases physical aggressiveness. However, the increase is almost always quite minor. *A number of investigators have argued that such a small effect is too weak to account for the major increase in rates of violence.* These investigators, however, overlook a key factor. Homicide is an extreme form of aggression. It's so extreme that only one person in 20,000 committed murder each year in the United States in the mid-1950's. If we were to rank everyone's degree of physical aggressiveness from the least aggressive to the most aggressive, the large majority of us would be somewhere in the middle, and murderers would be virtually off the chart. It is an intrinsic property of such "bell curve" distributions that small changes in the average imply major changes at the extremes. Thus, if exposure to television causes 8 percent of the population to shift from below-average aggression to above-average aggression, it follows that the homicide rate will double. The findings of the NBC study and the doubling of the homicide rate are two sides of the same coin.

Conclusion

In July of 2000, six major professional societies signed a joint statement on the hazards of exposing children to media violence. They were the American Psychological Association (APA), the American Academy of Pediatrics, the American Academy of Child and Adolescent Psychiatry, the American Medical Association, the American Academy of Family Physicians, and the American Psychiatric Association. They noted that "at this time, well over 1,000 studies point overwhelmingly to a causal connection between media violence and aggressive behavior in some children."

TV No Text

Television Violence Does Not Cause Real Violence

By those who believe that television violence is harmful, we have been told that there is overwhelming evidence that exposure to violence on television causes aggression. This claim is known as the causal hypothesis. We have been told that there is no longer any legitimate debate about this. We have also been told that the effect is as strong as the effect of cigarette smoking on cancer. We have been told that the press is biased because it gives more time to the opposing view than is warranted. Finally, we have been told that since there is no question, the press should not even mention the other view.

None of this is correct. The evidence is not overwhelming. Indeed, it provides no good reason to believe that television causes aggression, much less serious violence. The debate is certainly not over although some would like it to be. There is no comparison between the effect of smoking on cancer and the effect of television violence on aggression. And the press has, if anything, given far too much attention to the causal hypothesis than to those

who disagree with it. The most ardent advocates of the causal hypothesis seem to object to any disagreement or criticism of their position. However, their position is wrong and it deserves to be criticized. It is time once more to set the record straight.

What Does the Research Find?

First, most of the relevant work has shown that there is a correlation between exposure to violent television programming and aggression. That is, children who are exposed to more television violence tend to be more aggressive, and vice versa. The relationship is weak, but fairly consistent. However, the existence of this relationship does not indicate a causal relationship. *It provides no evidence that exposure to television violence causes children to be aggressive. The most likely explanation of the relationship is that some children are more aggressive in general than others, and that the more aggressive children prefer violent television, watch and play more aggressive games, and act more aggressively themselves.* To demonstrate that violent television causes aggressiveness, it is necessary to rule out this simple, intuitive explanation that is almost certainly at least partially true.

A review of all of the experimental research on both children and adults in 2002 showed that just about half of the studies were consistent with the causal hypothesis and half were not. When the studies that used the most dubious measures of aggression were eliminated, only 28 percent supported the hypothesis while 55 percent did not. The rest were either ambiguous or not directly relevant. When only studies involving children are considered, the findings are about the same but even weaker. There have been few relevant studies since 2002 so this summary is still accurate. *Thus, the main point about the experimental research is that overall, ignoring the methodological problems, the results do not support the notion that exposure to violent television makes children more aggressive.*

Problems With the Experiments

It is a mistake to take research at face value if it contains serious methodological problems. It has to be acknowledged that the experimental research on the effects of television violence is fraught with difficulties. There are three major problems.

First, the "violent" and "non-violent" programs are difficult if not impossible to equate. A basic principle of an experiment is that the two conditions one is comparing must be identical except for the key element. Suppose someone wanted to test the effect of a diet pill on weight loss with two groups of subjects. That person would not have the first group take the pill and eat only salads while the other group did not take the pill and could eat anything they wanted. If the first group lost more weight it could be due to the pill, but it could also be due to the differences in diet. So the right way to test this would be to give both groups the same diet and a pill, with the only difference being that one group took pills containing the diet medicine and the other group took pills with no medicine. Then if there were a difference in weight loss, it would be due to the only difference between the conditions, namely the content of the pills.

The same is true in the television violence research. The programs should ideally be identical except for the presence of violence in one program and the absence of it in the other. Unfortunately, this is very difficult to manage. Violence is typically such an integral part of the program that removing it leaves a program that makes no sense. There's no question that this is a difficult problem to solve.

Second, the measures of aggression are questionable. With a few exceptions, the studies do not measure the number of fights, amount of kicking or punching, or anything that most of us would consider aggressive behavior. Instead, they use analogues of real aggression that may or may not have much to do with the real thing. This is not easy to solve, because those who do the research cannot allow the participants to harm each other and generally cannot even allow them to play freely.

The third problem is that the children may be doing what they think the experimenter wants them to do or has given them permission to do. Participants in research often respond to what they perceive the researcher's intentions and values to be. *When a researcher shows children a violent program and then gives them the chance to be aggressive, the children may ask themselves why they were shown that program and what the experimenter has in mind.* A reasonable conclusion they could draw is that the researcher approves of the violence in the program and is, in essence, giving permission to behave aggressively.

This so-called demand factor is very powerful and can explain many of the findings. Since the effects of this demand factor are very well known, careful research is designed to avoid or at least to minimize it. If not, there is always the real possibility that any effects are due to demand. This is just standard research design. Yet those who have done the research on television violence do not seem to have been concerned by it and have largely ignored it.

Longitudinal Study

Perhaps the most widely mentioned study is what is now sometimes referred to as the 22-year study by Eron and Huesmann. This study first looked at children when they were 8, again 10 years later, and then again 22 years later when they were 30. This is a monumental piece of research that required vast effort and resources. Sadly, the findings have not been presented or discussed entirely objectively.

In terms of a causal effect of television violence on aggression, there was one result that provided good support for the causal hypothesis. From age 8 to age 18, there was a strong correlation between early exposure to violent television and later aggression. This correlation was stronger than the reverse correlation between later exposure and early aggression. However, this was true only for boys, not for girls, and on only one of the three measures of aggression. Yes, it is a nice finding, but it occurs in only one of six possible comparisons, which weakens it considerably.

In Huesmann's paper on the same participants when they were 30, he stressed one finding. The finding was that when early aggression is controlled statistically, early television viewing is related to the seriousness of criminal acts. That would indeed be impressive if it were not for the fact that there are many other results that are not consistent with this one or with the causal hypothesis. The study shows that early television viewing is not related to any measure of aggression in adult women, nor to any of six other measures of aggression in men. In other words, one out of 14 possible analyses supports the causal hypothesis. Since the authors do not correct their statistics as they should when doing multiple comparisons, the effect could well be due to chance.

With and Without Television

Several studies have looked at what happens when television is introduced into a country or community. If television causes aggression and violent behavior, assuming that the

television programming contains some violence, there should be more and more violent crimes after television is available. Centerwall (1989) compared the United States and Canada after the introduction of television with South Africa, which during the same period did not have television. The argument was that because the crime rate went up in the United States and Canada but not in South Africa, this showed that television caused crime.

As many have pointed out, this makes little sense since one cannot compare the situation in the three countries. In addition, other countries that also got television at the same time as the United States and Canada did not have the same increase in crime.

Another study (Williams, 1986) identified three small communities in Canada. One of these had several television stations, one had only one station, and the third had no television. It was known that television was about to be introduced into the third community. As a result, the researchers obtained measures of aggression in all three communities before and after the third community got television. The authors report that aggression went up in the third community compared to the others.

As with the comparison of countries, this study suffers from the insurmountable problem that the three communities differ in many ways that could explain the finding. Moreover, the actual results were far weaker than reported. They also failed to distinguish between real and play aggression, and were not consistent across measures. In any case, the television that was introduced into the third community was from the Canadian Broadcasting Company and had virtually no programs that contained any amount of violence. Thus, even if it could be shown that there was an effect, it could not have been due to violent programming.

Conclusion

Those who favor the causal hypothesis often say that although one type of research may not be conclusive, there is a confluence of findings such that all of the methods produce support. The opposite is true. That is, there is a confluence of findings such that all of the methods fail to produce support for the causal hypothesis. *Not one method has produced a clear majority of findings consistent with the idea that exposure to violent television makes people aggressive. It is not true for children and it is not true for adults.* Far from the overwhelming evidence that those who favor the hypothesis promote, the evidence overall and from each type of study is inconsistent, weak, and generally non-supportive.

In sum, there is no convincing scientific evidence that television violence causes children to be aggressive, or that any particular depiction of violence on television has this effect, or that it affects any particular type of children more than others. There has been a considerable amount of research on this topic, enough so that if there were an effect, the research should have shown it. Therefore, there is no effect of television violence on aggression. Even if there is an effect, it is vanishingly small because otherwise the research would have found it.