Here's Hoping it's not Just Text Structure:

The Role of Emotions in Knowledge Revision and the Backfire Effect

Neil G. Jacobson

University of Southern California

Ian Thacker

University of Texas, San Antonio

Gale M. Sinatra

University of Southern California

Author Note

Supported by an internal grant from USC Rossier School of Education.

Correspondence concerning this article should be addressed to Neil G. Jacobson, Rossier School of Education, ngjacobs@usc.edu

The Version of Record of this manuscript has been published and is available in Discourse Processes https://doi.org/10.1080/0163853X.2021.1925059

Abstract

Refutation texts are designed to facilitate the revision of inaccurate knowledge; however, studies have documented backfire effects wherein respondents become less accurate when exposed to a factual correction compared to another. Here, we explored whether epistemic emotions mediated knowledge revision or backfire processes when reading experimental refutation texts that varied by supporting information. We asked 294 online readers to report their knowledge and attitudes about genetically modified foods (GMFs) before randomly assigning them to one of three refutation text conditions which varied in type of supporting information. We documented relatively low knowledge (19.7%) and attitude (14%) backfire across conditions and found no evidence of backfire effects among types of supporting information. All texts facilitated knowledge revision regardless of the type of supporting information but did not differentially elicit epistemic emotions. For those who revised their inaccurate knowledge, hope and surprise mediated the knowledge revision process. However, those who demonstrated a backfire effect reported significantly more anger and confusion than readers who revised their inaccurate knowledge. Implications for research and practice are discussed.

Keywords: refutation text; backfire effect; conceptual change; belief change, knowledge revision; emotion

Here's Hoping it's not Just Text Structure:

The Role of Emotions in Knowledge Revision and the Backfire Effect

Adults encounter a stream of accurate and inaccurate information from online sources and social media platforms which require either integration with or revision of ones' prior knowledge and attitudes (Rapp & Braasch, 2014). These processes are fraught with obstacles including communicating and comprehending difficult scientific information that are exacerbated when the science is politicized. These difficulties are not simply about the information being communicated, but extend to how individuals feel while thinking (e.g., epistemic emotions) and may provide a more nuanced explanation of how the public's understanding of science is often at odds with scientific consensus views (Sinatra & Hofer, 2016). Given the importance of aligning the public and scientific community, researchers have developed techniques for revising readers' inaccurate knowledge specifically, refutation texts. However, many studies have ignored the role of epistemic emotions (e.g., surprise, confusion, anger; namely those experienced during knowledge generating activities) in addressing the backfire effect (Lewandowsky et al., 2012).

One technique that has been used extensively to revise inaccurate knowledge is refutation texts (Guzzetti, 2000; Guzzetti et al., 1993; Tippett, 2010). According to Kendeou et al. (2014) refutation texts typically contain three components which state the common but inaccurate knowledge, note that it is in fact incorrect, and provide information to support the accurate scientific knowledge. Refutation texts have been used to address inaccurate scientific knowledge for topics including force (Kendeou & van den Broek, 2007), seasonal change (Cordova et al., 2014; Will et al., 2019), and mathematics (Lem et al., 2017). Refutation texts have also been used to address social-scientific issues defined as social issues with conceptual or technological ties to science (Sadler, 2004). These social-scientific issues include vaccination, (Kessler et al., 2019; Vaughn & Johnson, 2018), genetic modification (Heddy et al., 2016; Trevors et al., 2016), and the greenhouse effect (Danielson et al., 2016).

Revising inaccurate knowledge about social-scientific issues is more difficult given readers likely have inaccurate knowledge and positive or negative attitudes that influence their thinking and may drive a *backfire effect* (Lewandowsky et al., 2012; Trevors et al., 2016). Lewandowsky et al. (2012) describe the backfire effect as the ironic strengthening of an original belief in information that is the subject of an attempted correction. Wood and Porter (2019), following Nyhan and Reifler (2010), operationalized the backfire effect as, "when the average respondent is made less accurate on a factual question when exposed to a false claim and its correction" (p. 136). Cook and Lewandowsky (2011) identify three backfire effects including *familiarity, worldview,* and *overkill.* Familiarity backfire describes the finding that more familiar information has a greater likelihood of being endorsed as true. Worldview backfire highlights the importance of "warm constructs" including prior attitudes, political identity, and personal values (Heddy et al., 2016; Muis et al., 2015; Trevors et al., 2016). The final backfire effect, overkill, is best described by a simple myth that is more cognitively attractive than an over-complicated correction (Cook & Lewandowsky, 2011). This backfire effect also highlights warm constructs, specifically the role of epistemic emotions (e.g., confusion, frustration, boredom, surprise, curiosity, anger) in studying how different types of supporting information (e.g., historical example, causal explanation, traditional explanation) impact knowledge revision and backfire processes (Arguel et al., 2019; Baker et al., 2010; D'Mello & Graesser, 2014; D'Mello et al., 2014; Lombardi et al., 2016; Silvia, 2010; Vogl et al., 2019). Therefore, we reframe the recommendation of Cook and Lewandowsky (2011) that, "It's not just *what* people think that matters, but *how* they think" (p.1), we argue it is also *how they feel while thinking* (Immordino-Yang & Damasio, 2007; Kaplan et al., 2016; Trevors et al., 2017).

Given concerns about inducing backfire effects, what should we tell readers when communicating knowledge that conflicts with their own, and relatedly should we attend to readers' emotions while reading refutation texts? Kendeou et al. (2013) and Kendeou et al. (2014) addressed the first question "What should we tell them?" by systematically investigating the role of the explanations in successful knowledge revision. In some instances, a simple refutation was adequate to facilitate revision, in others a single causal explanation was adequate. However, the greatest gains arose from the use of a refutation plus a three-sentence causal explanation. Interestingly, this conflicts with the recommendation of Cook and Lewandowsky (2011) who warn against providing overly detailed explanations (i.e., too much supporting information).

Returning to the second question "Should we attend to reader emotions?" recent evidence unequivocally says yes (Trevors & Kendeou, 2019; Trevors et al., 2017; Trevors et al., 2016). Trevors et al. (2016) provided evidence for a worldview backfire effect when learning about genetically modified foods (GMFs) by documenting how the importance of dietary-purity for an individual influenced their emotions while reading a refutation text. Also, Trevors et al. (2017) documented that readers' emotions differed while reading refutation texts of similar structure, but with different combinations of refutation text components and supporting information. Given the variation in findings about knowledge revision and emotions, Bohn-Gettler (2019) proposed a framework to disentangle findings about emotions by specifying the learning process, here we are interested in knowledge and attitude revision and backfire effect processes.

In this investigation we examined two backfire effects: overkill and worldview. Overkill backfire was manipulated via the supporting information paired with the refutation text, and worldview backfire was assessed by measuring attitudes toward GMFs before and after reading. By assessing knowledge and attitudes, we could examine whether individuals demonstrated a knowledge and/or an attitudinal backfire effect. Overkill backfire was assessed by manipulating the amount of information provided to support the accurate knowledge (Kendeou et al., 2013; Kendeou et al., 2014; Rapp & Kendeou, 2007, 2009). Key to our investigation is how refutation texts and their supporting information impacted readers' epistemic emotions and if these individual emotions (e.g., confusion, surprise, anger) mediated knowledge and/or attitude revision or backfire effect processes when learning about GMFs.

To motivate our investigation, we first describe the refutation text structure highlighting the functions served by supporting information. Next, we describe the Knowledge Revision Components Framework (KReC; Kendeou & O'Brien, 2014) relative to the refutation text structure attending to the role of supporting information in knowledge revision. Finally, we describe research on warm constructs including attitudes and emotions arguing that supporting information may activate epistemic emotions that mediate the backfire and knowledge revision processes.

The Refutation Text

The refutation text is typically defined as having a three-part structure (Guzzetti et al., 1993; Tippett, 2010). Kendeou et al. (2014) state the text, "must present previously acquired but no longer correct information; it must state that the information is, in fact, incorrect; and it must contain evidence to support the correct to be learned information" (p. 376). We agree with the description of the first two components—that it is important to present inaccurate information and provide a refutation cue—however**,** we describe the third component (that refutations must contain supporting evidence) as two component pieces: the correct to-be-learned information and information to support the to-be-learned information.

Generally, explanations induce dissatisfaction (Posner et al., 1982), highlight contradiction (Thagard, 1989), privilege a subset of beliefs that constrain reasoning (Lombrozo, 2006), or provide a more integrated network of information (Kendeou et al., 2013; Kendeou et al., 2014). Importantly, these functions describe causal explanations, with research indicating that increased causal explanatory information produces the greatest learning gains. However, the overkill backfire effect (Cook & Lewandowsky, 2011) and a recent meta-analysis (Chan et al., 2017) warn against overly detailed explanations (such as explanatory information). Based on a meta-analysis across 52 studies (N = 6,878), Chan et al. (2017) found that studies coded as providing “highly detailed” information when debunking a misconception were associated with larger knowledge revision effects compared with “less detailed” information. However, Chan et al. (2017) also found that more supporting information (i.e., message detail) correlated with misinformation persistence. Similarly, simple explanations that contain fewer causal explanations are preferred to overly detailed explanations (Lombrozo, 2006, 2007). A final note on explanations comes from Brem and Rips (2000) who state, “there is not a single best form an argument takes and an arguers tactics cannot be evaluated on a single unchanging set of criteria” (p. 578).

We highlight this methodological difference in text structure as a possible mechanism for the overkill backfire effect. Prior work on knowledge revision shows variations in supporting information including causal explanations (Rapp & Kendeou, 2007, 2009), attitudinal statements (Thacker et al., 2020), and emotional content (Trevors & Kendeou, 2019) influence knowledge revision. Given this variation, we argue that different types of supporting information may activate different emotions (e.g., confusion, curiosity, surprise) or the same emotions differently (i.e., same emotion, but at a different intensity) and may provide a more nuanced account of the role of emotions in understanding backfire and revision processes.

The Knowledge Revision Components (KReC) Framework and Refutation Texts

Kendeou and O'Brien (2014) introduced the KReC, a model of knowledge revision that articulates two assumptions regarding the human information processing system and three principles specifying how knowledge revision can occur. Both are directly applicable to learning from refutation texts and understanding the knowledge revision and backfire processes. The two assumptions are encoding and passive activation while the three principles are *co-activation*, *integration*, and *competing activation*. Foundationally, the encoding principle assumes that information cannot be erased or replaced in memory, although interference and decay can. Additionally, given that memory is associative and passive activation of long-term memory is acknowledged in the model there is a mechanism whereby prior knowledge, attitudes, and emotions toward a topic can be activated when learning about a controversial topic (Broughton et al., 2011; Ecker et al., 2014; Heddy et al., 2016; Sinatra & Seyranian, 2016).

Based on these assumptions, the first principle, co-activation, extends from the passive activation assumption and provides the mechanism by which accurate and inaccurate knowledge are co-activated in a reader's memory, a necessary condition for knowledge revision (Kendeou & van den Broek, 2007). The second principle, integration, articulates how the newly encoded information, whether accurate or not, must be integrated with prior knowledge for revision to occur. The third and final principle, competing activation, describes the iterative process by which knowledge (accurate or inaccurate) is organized and integrated in a readers' long-term memory. Competing activation also provides the rationale for how and why supporting information facilitates knowledge revision by providing a rich network of integrated information.

KReC principles are relevant to the refutation text structure and our experimental manipulations. The co-activation principle is seen in the text structure by activating prior knowledge and alerting readers that incoming information may conflict with their knowledge before stating the accurate scientific knowledge. This process facilitates co-activation; however, it is inadequate for knowledge revision to occur. For revision to occur, integration of the accurate and inaccurate knowledge must occur in long-term memory. For example, the knowledge that GMFs are not clones must become integrated in a way that allows for competition.

The requirement for integration into long-term memory relates directly to the final principle: competing activation, which provides the mechanism for knowledge revision and is key to our investigation of supporting information. In a series of studies Rapp and Kendeou (2007, 2009) and Kendeou and colleagues (2013, 2014) investigated the importance of each refutation text component on knowledge revision. Overall, results indicated that it is both the quality of supporting information (i.e., causal explanations) provided to a reader and the amount of said information that produce the refutation text effect. For example, the authors found that while a one-sentence causal explanation was adequate to eliminate disruption, only the three-sentence causal explanation stopped reactivation of inaccurate knowledge. Interestingly however, some evidence emerged for the efficacy of the explanation or the refutation alone although the effects did not produce differences in learning outcomes.

Together these principles make clear the importance of the three-part refutation text structure, specifically activating the inaccurate knowledge, providing a cue to alert readers to an incoming conflict, and stating the accurate knowledge. However, it also makes apparent the gap between the prescribed three-part structure and our argument for the importance of the fourth component, the supporting information. In the following section, we review research on warm constructs (Pintrich et al., 1993; Sinatra, 2005) including attitudes and epistemic emotions and their relationship to supporting information when learning about social-scientific issues.

Warm Constructs and Supporting Information

Warm constructs refer to motivational variables including attitudes, and emotions (Pintrich et al., 1993; Sinatra, 2005). An illustrative example is the demotion of Pluto to dwarf planetary status, which when viewed through a warm constructs lens foregrounds the role of knowledge, attitudes, and emotions in science learning (Broughton et al., 2011). Returning to the worldview backfire effect which addresses the role of attitudes and emotions in the processing of scientific messages we first define and then review relevant findings in the area.

Attitudes. Attitudes are positive or negative evaluations of an object, event, or person (Eagly & Chaiken, 1993) which can impact how individuals engage with a science topic and can influence their interpretation of the science (Heddy et al., 2016; Sinatra & Seyranian, 2016). Sinatra and Seyranian (2016) extend our understanding of attitudes with, "the idea that attitudes can be informed by one or two types of beliefs, [inaccurate and accurate], which can operate alone or simultaneously to inform evaluations" (p. 247). Prior research has indicated that a reader may hold the inaccurate knowledge that GMFs are made by injecting hormones which contributes to their negative attitude about GMFs (Heddy et al., 2016; Thacker et al., 2020). According to this perspective, attitudes and knowledge are inherently linked and rarely shift in isolation (Heddy et al., 2016; Sinatra, Kienhues, et al., 2014). Thacker et al. (2020) empirically tested this perspective documenting that knowledge and attitudes were dependent and best modelled as such. Finally, when revision does occur, increasing evidence points to the role of emotions in the process (Thacker et al., 2020; Trevors & Kendeou, 2019; Trevors et al., 2017; Trevors et al., 2016).

Emotions. Research has shown that how we think and learn are inexorably linked to our emotions (Pekrun & Stephens, 2012; Sinatra, Broughton, et al., 2014). Immordino-Yang and Damasio (2007) argue that when emotions are integrated into human decision-making a new conceptualization emerges: *emotional thought.* This view allows us to foreground how individual emotions may serve as an "emotional rudder" that guides learning. Here, we align with Bohn-Gettler (2019), who defines emotions as, "transient, state based emotions that a reader may experience at a particular moment in a specific situation" (p. 389). This view allows for emotions to be classified along two dimensions: valence and activation (Pekrun & Stephens, 2012; Pekrun et al., 2017). Valence refers to whether an emotion is experienced as positive or negative, such as enjoyment versus anger. Activation describes the intensity associated with the emotional experience which varies from low to high.

In addition to valence and activation, emotions can arise from and be focused on different objects. Epistemic emotions arise from knowledge and knowledge generating activities (Brun & Kuenzle, 2008) such as resolving conflicting information (Pekrun & Stephens, 2012; Pekrun et al., 2017). Epistemic emotions include surprise, interest, curiosity, enjoyment, happiness, confusion, anxiety, frustration, and boredom. Additionally, research on GMFs has identified anger, hopefulness, and hopelessness as important emotions in knowledge revision for this topic (Heddy et al., 2016; Lombardi et al., 2016; Thacker et al., 2020; Trevors et al., 2016).

Refutation texts are designed to trigger conflict and ideally promote knowledge building activities; therefore, it is vital to understand how readers experience the supporting information provided in the text and how these emotions mediate knowledge revision. One benefit of the refutation structure is the high level of coherence it provides by featuring overlap between and among scientific information (Sinatra & Broughton, 2011). However, McNamara (2001) argues that a less coherent text facilitates processing by forcing the reader to process the text more actively, an argument that less supporting information may be beneficial (D'Mello et al., 2014; Vogl et al., 2019). Conversely, a less coherent text may cause confusion, and if unresolved, may lead to frustration or boredom that inhibits learning (Baker et al., 2010; D'Mello et al., 2014). Readers may attempt to resolve their confusion by leveraging prior knowledge and attitudes to resolve the discrepancy. Refutation texts are thought to highlight conflict between prior knowledge and incoming information by triggering surprise that can prompt interest, curiosity, and confusion which can then lead to amelioration of misconceptions (D'Mello et al., 2014; Munnich & Ranney, 2019; Ranney et al., 2016; Thagard, 1989; Vogl et al., 2019).

Ranney and colleagues (2019) found evidence that the intensity of surprise brings about different types of processing. If surprise is too low, new information can be assimilated into an existing network of inaccurate knowledge, whereas if surprise is experienced more intensely it can induce two types of processing accommodation (knowledge revision) or rejection (a backfire effect; Munnich & Ranney, 2019). Finally, explanations are argued to provide a richer more integrated network that facilitates competing activation and supports knowledge revision (Kendeou & O’Brien, 2014); however, this conflicts with Cook and Lewandowsky’s (2011) warning to convey simple explanations as they are generally preferred (Lombrozo, 2007). These conflicting findings make it unclear how readers would emotionally experience refutation texts paired with various supporting information.

The Present Investigation

The purpose of the present study was to examine the role of supporting information used to communicate scientific knowledge and how readers emotionally experience this information. While supporting information (i.e., a causal explanation) is a driving force in knowledge revision (Kendeou et al., 2013; Kendeou et al., 2014; Rapp & Kendeou, 2009), it may also be a driving force in backfire effects (Chan et al., 2017; Cook & Lewandowsky, 2011).

We hypothesized that all readers would benefit from the refutation texts as they adhere to the traditional three-part structure and only vary on the amount of supporting information provided (Guzzetti, 2000; Kendeou et al., 2013; Tippett, 2010). The "traditional" condition featured only a single statement of the accurate scientific knowledge. The "causal explanation" condition featured several sentences providing a causal explanation of the scientific knowledge. This condition required us to create a text comparable in length but without a causal explanation. To accomplish this, we created an "example" condition that provided an example of the accurate knowledge such as discussing the potato famine in the context of cloning.

We hypothesized that those who read the causal explanation would experience the greatest knowledge revision from pre to post, followed by those who read the example condition, with the least revision occurring for those who read the traditional explanation condition. Our rationale was based on prior research that found causal explanations best facilitated knowledge revision (Kendeou et al., 2013; Kendeou et al., 2014) and that more, compared to less supporting information facilitated revision (Chan et al., 2017). Recall however, Chan et al. (2017) also found that misinformation persisted when more supporting information (i.e., message detail) was presented. This led us to ask whether knowledge revision in refutation texts is driven by the supporting information (i.e., traditional, causal, example), or is merely providing supporting information adequate to facilitate revision and avoid backfire?

We hypothesized that backfire effects would be most evident in the traditional condition because participants did not receive supporting information (either causal or example) for the scientific knowledge which may induce negative epistemic emotions such as confusion, frustration, boredom, and anger. Conversely, we posited that epistemic emotions would play the greatest role in the causal explanation condition with low negative emotions (i.e., confusion, frustration, boredom, and anger) and high positive emotions (i.e., interest, surprise, curiosity) followed by the example condition which may induce greater surprise or interest which is associated with knowledge revision (Munnich & Ranney, 2019).

Our specific research questions were,

1. Do readers shift their knowledge and attitudes pre to post after reading refutation texts with different supporting information?
2. Do readers differentially backfire (in knowledge and/or attitudes) dependent on the supporting information provided?
3. Do epistemic emotions (e.g., surprise, anger, confusion, curiosity, interest) mediate the knowledge revision and backfire processes?

Methods

Participants

Participants (*N* = 328) were recruited and compensated for participation using Amazon's Mechanical Turk with participation criteria including living in the US and speaking English.[1] To exclude participants who may have devoted little attention to the survey or completed the survey over the course of several sittings, we dropped the fastest and slowest 5% from analyses as a

[1]The sample size of approximately 300 (100 participants in each of 3 text) was a rounded estimate based on power analyses for a 1x3 ANOVA to detect an effect size of .1 with power of .9 using the "pwr2" package in R (Lu, Liu, & Koestler, 2017). This sample is also acceptable for detecting mediation effects (Fritz & MacKinnon, 2007).

means of quality control (Kees et al., 2017) leaving *N* = 294 observations in the analytical sample.[2] After dropping these participants, reading time was a mean of 86.5 seconds (*SD* = 52.6 seconds) and median of 78 seconds. A data collection error resulted in only collecting 62% (183 of 294) of reader demographics. Demographics on the partial sample revealed that participants' average age was 41.2 (*SD* = 12.9) and ranged from 21 to 76. Participants were majority female (59%) and self-reported their ethnicities as follows: White (81%), African American/Black (6%), American Indian/Alaskan (0.5 %), Mexican American/Chicano (1.1 %), Asian (4.9%), Hawaiian/Pacific Islander (0.5 %), other (2.2 %) with 4.9% reporting more than one ethnicity.

Materials

Experimental refutation texts. Refutation texts were designed to refute reader misconceptions by, (a) activating the misconception, (b) providing a cue that this conception is incorrect, and (c) stating the accurate scientific knowledge. A single 377-word refutation text was used to target four misconceptions about GMFs and has been used in previous research (Heddy et al., 2016). The misconceptions include the inaccurate knowledge that, (a) hormone injection is genetic modification (Uzogara, 2000), (b) cloning is genetic modification (Gaskell et al., 1999), (c) genetic modification can only occur in laboratories (Uzogara, 2000), and (d) genetic modification is only a contemporary process.

This single 377-word text served as the traditional refutation text from which two additional texts were created. An example from the traditional text is, "You may think that genetically modifying foods is the same process as cloning. This belief is not correct. Cloning involves making an exact genetic copy of an organism. All of the genetic information is identical

[2] Note that we ran all analyses with the full sample. Though coefficients were slightly different, the significance levels and direction of effects for the main findings were identical.

between those two organisms." From this traditional text, a 571-word causal explanation text was created. This text provided a causal explanation for each misconception, for example, "In contrast, the process of genetically modifying food can be done using gene cloning methods; however, the protein in the genetically modified organism has been modified somewhat so that the host (modified) organism will express the desired trait. Thus, the genetically modified organism is not usually an exact replica of the donor organism." A final 541-word example text was created to approximate the length and structure of the causal explanation text. This text provided an historical example such as, "In 1856 Gregor Mendel, a monk and teacher, began studying cross-pollination in peas and this work provided the basis for our modern understanding of genetics." Across all three texts, other than word, and sentence counts the texts were roughly equivalent for grade level (traditional = 10.9, causal = 11.3, example = 11.3). Finally, these texts were not designed to influence attitudes toward GMFs but instead aimed to present the scientific evidence only.

Genetically modified food knowledge measure. A 10-item multiple choice knowledge measure about GMFs was completed by participants before and after exposure to the experimental manipulation. Participants were asked to select the response option that aligned with their knowledge and has been used in previous investigations (Heddy et al., 2016; Thacker et al., 2020; Trevors et al., 2016). For each question response options included the accurate knowledge in addition to three distractors presenting common misconceptions about GMFs. The measure of knowledge ranged from 0-10 and corresponded to the number of correct responses a reader provided. Cronbach's alpha was .68 for pre-intervention knowledge and .81 for post-intervention knowledge. Following Wood and Porter (2019), we created a binary variable to indicate "knowledge backfire" with individuals whose mean knowledge score decreased from pretest to posttest (coded as 1) and increased or stayed the same (coded as 0).

Attitudes toward genetically modified foods. A 5-item Likert measure of GMF attitudes was administered before and after exposure to experimental condition. Participants indicated their agreement on a 5-point Likert scale anchored at strongly disagree (1) and strongly agree (5) modified from (Lorenzoni et al., 2006) to fit our topic. For example, an item stated, "I approve of genetically modified foods" with higher scores indicating a more positive attitude. Cronbach's alpha was .95 for pretest attitudes and .96 for posttest attitudes. Again, we created a binary variable to indicate "attitudinal backfire" with individuals whose mean attitude became more negative from pretest to posttest (coded as 1) and positive or stayed the same (coded as 0).

Epistemic and Epistemically Related Emotions.

We measured readers' epistemic and epistemically related emotions using a modified form of the Epistemic Emotions Scale (EES; Pekrun et al., 2017), which are those experienced during knowledge generating activities. Immediately after reading we asked readers to indicate on a 5-point Likert scale (1 = *not at all* to 5 = *very strong*) how much they experienced the following 13 emotions: confused, frustrated, anxious, hopeless, angry, hopeful, curious, interested, happy, fearful, bored, surprised, and enjoyment. Individual emotion ratings were used in analyses and treated as continuous variables.

Procedure

A study description was posted to Amazon's Mechanical Turk wherein interested participants could follow a link to our Qualtrics survey. After providing informed consent participants were randomly assigned to one of three between-subjects text conditions (traditional, causal, or example). All participants were measured at two time points, resulting in a 2-Time (pre vs post) x 3-Text Condition (traditional vs. causal vs. example) mixed design.

First, participants completed the knowledge and attitude measures. Second, participants were randomly assigned to one of three refutation text conditions, where they could read and advance the screen at their own pace. After reading, participants were asked to complete the epistemic emotions measure, knowledge and attitude measures, and a demographic data sheet. After completing the session, participants were debriefed on the purpose of the study, provided a Mechanical Turk payment code, and thanked for their time.

Results

Preliminary Analyses

All variables were examined for skewness and kurtosis. Skewness ranged from -.52 to 1.87 and kurtosis ranged from -1.43 to 2.65, which are both acceptable (Tabachnick & Fidell, 2013). To examine equivalence at pretest, we tested for differences in prior knowledge and attitude across conditions documenting no significant differences for prior knowledge (p = .853) or attitudes (p = .963). Raw means and standard deviations overall and by outcome (i.e., knowledge revision, knowledge backfire, attitude revision, and attitude backfire) are presented in Table 1. Additionally, a Chi-square test revealed that knowledge and attitudinal backfire were not independent (χ^2 = 7.33, p = .0068), with frequencies that were higher than expected among those who backfired in both knowledge and attitude (χ^2 contribution = 6.05), providing evidence for

Sinatra and Seyranian's (2016) knowledge-attitude link framework. Finally, correlations among all variables of interest are shown in Table 2.

Differences in Knowledge and Attitude Revision by Supporting Information

To answer our first research question, *Do readers shift their knowledge pre to post after reading a refutation text with different supporting information?* we ran two two-way mixed design ANOVAs, one for mean knowledge as the main outcome, the other with attitude as the main outcome. For both, the within variable was time and the between variable was text condition. For knowledge, we found a significant main effect of time $F(1, 291) = 120.07$, $p < .001$, η*-squared* = .41 (i.e., readers' accurate knowledge increased from pre to post) but no main effect of text condition $F(2, 291) = .137$, $p = .872$, or their interaction $F(2, 291) = .591$, $p = .555$. The same was true for attitudes; we found a significant main effect of time $F(1, 288) = 37.06$, $p <$. 001 (i.e., readers' attitudes became more positive from pre to post), η*-squared* = .13, no effect of text condition, $F(2, 288) = .007$, $p = .993$ and no significant interaction $F(2, 288) = .355$, $p = .702$. That is, there was evidence of knowledge correction over time, but no differences based on the supporting information.

Knowledge, Attitude, and the Backfire Effect

To answer our second research question, *Do readers differentially backfire* (in knowledge and/or attitudes) *dependent on the supporting information provided?* we first ran a Chi-squared analysis to test differences in proportions of readers who demonstrated backfire in knowledge and/or attitude between the three supporting information conditions. Regarding knowledge backfire, 19.7% backfired overall, 15% in the example condition, 22% in the traditional condition, and 21% in the causal explanation condition. The Chi-squared test revealed no significant differences in proportions by condition $\chi^2(2, 294) = 1.75$, $p = .41$. Similarly, we found that few (14%) participants backfired in attitude overall, 13% in the example condition, 16% in the traditional condition, and 15% in the causal condition and no significant dependence between text condition and attitudinal backfire ($\chi^2(2, 294) = 0.25$, $p = .88$). That is, we found no evidence that readers' backfire rates were dependent on the type of information provided.

Differential Rates of Knowledge and Attitudinal Backfire by Supporting Information

To observe whether additional background characteristics predicted backfire, we conducted logistic regressions with knowledge and attitudinal backfire as the main outcome variables. Table 3 summarizes the logodds regression coefficients when knowledge and attitude backfire was the main outcome measure. For each outcome, we reported three models: one included the text conditions as predictors (with the "traditional" condition as the reference group), the second also included knowledge, attitudes, and their interaction, and the third included these predictors along with each of the 13 epistemic and epistemically related emotions. This model building approach accords with our framing of learning controversial science content. The first model is "cold" as it only accounts for readers' knowledge while the second and third models incorporate "warm" constructs. The second model builds on our finding on nonindependence between knowledge and attitudes in line with Thacker et al. (2019) and Sinatra and Seyranian (2016) and assesses the possible relationships between knowledge and attitude (i.e., main effects of either, each, or their interaction). However, only the third model incorporates epistemic and epistemically related emotions as a measure of "how readers feel while thinking"—a preliminary step in understanding the role of emotional rudders in knowledge revision.

Knowledge backfire. When knowledge backfire was the main outcome, the first model did not fit significantly better than the null $\chi^2(2, 294) = 1.81$, $p = .40$), but improved after including knowledge, attitude, and their interaction (χ^2 (5, 294) = 16.23, p = .006) and improved further after adding the 13 epistemic emotions (χ^2(18) = 51.9, p < .001). We found again that, after controlling for all other variables in the model, supporting information was not a significant predictor of backfire (for all three models, Wald test-statistic < 3.0, p > .22). Including prior knowledge and attitude in model two, we found a significant positive effect of prior knowledge (b = .77, p = .002) and a marginally significant effect of prior attitude (b = .604, p = .07) indicating that prior knowledge and attitudes each positively related to knowledge backfire—effects that may be due to participants with exceptionally high prior knowledge and attitudes regressing to the mean at post-test. However, we found a significant interaction between prior knowledge and attitudes (b = -.17, p = .012) in predicting knowledge backfire logodds. That is, people with greater knowledge *and* positive attitudes were less likely to demonstrate knowledge backfire than other groups providing evidence for the simultaneous modeling of knowledge and attitudes in our later analyses.

Including each epistemic and epistemically related emotion in model three, only anger was a significant and positive predictor of knowledge backfire (b = .563, p = .030). No other emotion significantly predicted knowledge backfire (all p > .29). The knowledge and attitude interaction remained significant (b = -.189) as did the significant effect of knowledge and the marginally significant effect of attitudes.

Attitude backfire. Similar analyses were conducted with attitude backfire as the main outcome variable (coefficients are summarized in Table 3). The model including supporting information alone was not better fitting than the null (χ^2 (2, 294) = 0.26, p = .88), which improved after including knowledge, attitude, and their interaction (χ^2 (5, 294) = 15.50, p = .008) and improved further after adding epistemic emotions (χ^2 (18) = 39.1, p = .003). We found again that, after controlling for all other variables in the model, supporting information was not a significant predictor of attitudinal backfire (for all three models, Wald test-statistic < 0.47, $p > .79$). Including prior attitude and knowledge in model two, we found a significant positive effect of prior attitude ($b = .69$, $p = .041$) and a marginally significant interaction between prior attitude and knowledge ($b = -.16$, $p = .078$) in predicting attitude backfire logodds.

Including each epistemically related emotion in model three, only confusion was a significant positive predictor of attitudinal backfire ($b = .454$, $p = .041$). No other emotion significantly predicted backfire (all $p > .08$). The effect of prior attitude remained significant ($b = .91$, $p = .022$), as well as the marginally significant interaction between attitude and knowledge ($b = -.164$, $p = .095$).

Path Models

To answer our third question, *Which emotions mediate the knowledge revision and backfire processes?* we explored the relationships between and among knowledge, attitudes, and epistemically related emotions to understand their relationships with successful revision and unsuccessful backfire. We modeled the relationship using discrete (i.e., individual) emotions as mediators of knowledge-attitude relations described in Thacker et al. (2020) as opposed to modeling positive and negative emotions with latent variables (Heddy et al., 2016), as we were interested in the influence of specific emotions on knowledge revision.[3]

[3] Note that we reran all analyses with positive emotion and negative emotion composite variables and replicated the main findings regarding interactions between prior attitudes and knowledge in the logistic regressions when the outcome was either knowledge or attitude backfire. Similarly, the main findings from the path analyses were replicated. Namely, we found that for the full sample, positive emotions significantly mediated relations between pre-reading and post-reading attitudes, and negative emotions mediated relations between pre-reading attitudes and post-reading knowledge and attitudes. This was also the case for those who experienced knowledge revision, but no significant mediations for those who backfired in knowledge, and identical results when the sample was split by attitude backfire.

We first ran the model recommended by Thacker et al. (2020) with the full sample to test for relationships between attitude and knowledge, and for significant indirect effects of emotions (see Figure 1 for the hypothesized model). We ran the model four more times, to systematically explore the relationships documented by Thacker et al. (2020). First, we analyzed knowledge, limiting the sample to those who demonstrated knowledge revision. Next, we analyzed those who demonstrated knowledge backfire. Third, we analyzed attitudes beginning with successful revision followed by those who demonstrated attitudinal backfire. All path models were constructed using Lavaan 0.6-3 (Rosseel, 2012).

Full sample. For the full sample, the model resulted in acceptable fit at conventional levels ($\chi^2(df = 45$, n = 286) = 181.089, *CFI* = .954, *RMSEA* = .103; e.g., Hu & Bentler, 1999). For a summary of the coefficients found for this analysis, see Figure 2. As hypothesized, we found that prior knowledge significantly and positively predicted prior attitudes (b = .193, SE = .03, p < .001). Pre-intervention attitudes then positively predicted hope (b = .498, SE = .05, p <.001), boredom (b = .155, SE = .05, p =.001), happiness (b = .199, SE = .05, p <.001), and hopelessness (b = .072, SE = .04, p =.039), and negatively predicted anger (b = -.336, SE = .05, p < .001), anxiety (b = -.099, SE = .04, p = 023), surprise (b = -.284, SE = .07, p < .001), frustration (b = -.108, SE = .04, p = .003).

Subsequently, epistemic emotions predicted post-intervention knowledge and attitude. Post intervention knowledge was positively predicted by surprise (b = .391, SE = .11, p < .001) and negatively predicted by hopelessness (b = -.636, SE = .11, p = .001). Moreover, surprise significantly and negatively mediated relations between pre-intervention attitude and knowledge (*indirect effect* = -.111, SE = .04, p =.006). Post-intervention attitude was significantly and positively predicted by hope (b = .132, SE = .04, p =.001), curiosity (b = .128, SE = .04, p

=.001), and surprise (*b* = .092, *SE* = .03, *p* =.002) and negatively predicted by fear (*b* = -.123, *SE* = .05, *p* = .011). Moreover, surprise significantly mediated relations between pre- and post-attitudes (*indirect effect* = -.026, *SE* = .02, *p* = .015) as did hope (*indirect effect* = .021, *SE* = .02, *p* = .001).

We then ran the same model separately for those who demonstrated knowledge revision (see Figure 3.1) and those who demonstrated knowledge backfire (Figure 3.2) to investigate whether emotions operated differently as mediators between the two processes (see Figure 3 for a summary of coefficients).

Readers who experienced knowledge revision. The model (Figure 3.1), when conducted with people who revised their knowledge had acceptable fit (χ^2 (*df* = 45, *n* = 236) = 136, *CFI* = .963, *RMSEA* = .092). Again, prior knowledge significantly and positively predicted attitudes (*b* = .243, *SE* = .03, *p* < .001). Pre-intervention attitudes then positively predicted hope (*b* = .504, *SE* = .06, *p* < .001), boredom (*b* = .149 , *SE* = .05, *p* = .003), and happiness (*b* = .193, *SE* = .06, *p* = .001), and negatively predicted fear (*b* = -.094, *SE* = .04, *p* = .033), anxiety (*b* = -.128, *SE* = .05, *p* = .007), surprise (*b* = -.318, *SE* = .08, *p* < .001), and anger (*b* = -.320, *SE* = .05, *p* < .001).

Epistemically related emotions then significantly predicted knowledge and attitudes. Specifically, post-knowledge was positively predicted by surprise (*b* = .420, *SE* = .10, *p* < .001) and negatively by hopelessness (*b* = -.533, *SE* = .18, *p* = .004). Surprise was also a significant mediator of the relationship between prior attitudes and final knowledge (*indirect effect* = -.134, *SE* = .045, *p* = .003). Post-attitudes were also significantly predicted by emotions, positively by hope (*b* = .140, *SE* = .05, *p* = .002), curiosity (*b* = .173, *SE* = .04, *p* < .001), and surprise (*b* = .093, *SE* = .03, *p* = .006) and negatively by confusion (*b* = -.081, *SE* = .04, *p* = .049). We tested for indirect effects of pre and post attitudes, and found that surprise (*indirect effect* = -.029, *SE* = .01, *p* = .021), and hope (*indirect effect* = .071, *SE* = .02, *p* = .003) were significant negative and positive mediators respectively.

Readers who demonstrated knowledge backfire. Using the same model but with readers who backfired in knowledge (see Figure 3.2), we found fewer relationships overall. The model fit was acceptable according to some test statistics χ^2 (*df* = 45, *n* = 50) = 102, *CFI* = .913), but not others (*RMSEA* = .159). One distinguishing characteristic of readers who backfired in knowledge is that prior knowledge did not significantly predict prior attitudes (*b* = .053, *SE* = .06, *p* = .365). Prior attitudes positively predicted hope (*b* = .467, *SE* = .149, *p* = .002), and negatively predicted confusion (*b* = -.209, *SE* = .09, *p* = .030), and frustration (*b* = -.337, *SE* = .11, *p* = .003). However, after the intervention, only anxiety (*b* = .261, *SE* = .12, *p* = .025) and hope (*b* = -.192, *SE* = .07, *p* = .009) were significant predictors of knowledge while no other emotion significantly predicted post-attitudes (all *p* > .068). We found no statistically significant indirect effects of emotions on relations between pre-reading attitudes and post-reading knowledge or attitudes.

Readers who demonstrated attitude revision. We used the same model for those who revised their attitudes (see Figure 4 for a summary of coefficients). We found that the model had acceptable fit when limiting the sample to readers who experienced attitude revision (χ^2 (*df* = 45, *n* = 244) = 169, *CFI* = .953, *RMSEA* = .106). Among this subsample, prior knowledge significantly and positively predicted pre-reading attitudes (*b* = .215, *SE* = .03, *p* < .001). Pre-intervention attitudes subsequently predicted a host of emotions including negative relations with anger (*b* = -.373, *SE* = .05, *p* < .001), anxiety (*b* = -.100, *SE* = .05, *p* = .035), surprise (*b* = -.293, *SE* = .08, *p* < .001), and frustration (*b* = -.115, *SE* = .04, *p* = .002) and positive relations with boredom (b = .149 , SE = .05, p = .003), happiness (b = .186, SE = .06, p = .001), and hope (b = .473, SE = .06, p < .001).

Epistemically related emotions then significantly predicted knowledge and attitudes. Specifically, post-knowledge was positively predicted by surprise (b = .363, SE = .11, p = .001). Surprise was also a significant mediator of the relationship between prior attitudes and final knowledge (*indirect effect* = -.106, SE = .05, p = .011). Post-reading attitudes were also significantly predicted by emotions, specifically positively by hope (b = .142, SE = .04, p < .001) and surprise (b = .063, SE = .03, p = .021) and negatively by fear (b = -.103, SE = .05, p = .031). We tested for indirect effects of pre and post attitudes, and found that surprise (*indirect effect* = -.018, SE = .009, p = .049), and hope (*indirect effect* = .067, SE = .02, p = .001) were significant negative and positive mediators respectively.

Readers who demonstrated attitude backfire. We used the same model but with readers who backfired in attitude, however the model did not have acceptable fit at conventional levels (χ^2 (df = 45, n = 42) = 99, CFI = .883, $RMSEA$ = .169). Therefore, we refrain from reporting results due to poor fit and lack of statistical power.

Discussion

Daily individuals seek and find scientific information and misinformation which elicit emotions likely to impact further processing. The proliferation of misinformation online and via social media platforms exacerbates the spread of scientific misconceptions. Refutation texts are designed to combat inaccurate information by confronting misconceptions and explaining correct knowledge. However, concerns about their use have increased since evidence exists that corrective efforts can cause readers to backfire (Chan et al., 2017; Nyhan & Reifler, 2010).

We tested whether variations in the supporting information used in a refutation text ameliorated or exacerbated knowledge and/or attitudinal backfire effects. Our results showed that supporting information did not, in this instance, have a differential impact on the likelihood of backfire or the epistemic emotions readers experienced. Overall, our findings tended to demonstrate the refutation text effect (i.e., a reduction in misconceptions from pre to posttest) regardless of refutation text structure manipulations. This suggests that the refutation text effect is robust, since our conditions were designed to “bend,” if not “break,” the effect by reducing the amount of supporting information readers were provided to facilitate their knowledge revision. We anticipated differential impacts of the refutation texts according to the nature of the supporting information (i.e., more surprise and interest and less confusion and frustration in the causal explanation condition). However, we found that causal statements, examples, or merely stating the accurate knowledge all facilitated knowledge revision, a finding that contributes to the growing body of refutation text modification studies (Kendeou et al., 2013; Schmid & Betsch, 2019; Thacker et al., 2020; Trevors & Kendeou, 2019).

Our findings did indicate some evidence of backfire (19.7% knowledge and 14% attitudinal), although we did not find evidence of differential backfire due to our experimental manipulations. Given the overall success of the refutation texts we focused on reader characteristics to examine the knowledge revision and backfire processes. In line with Sinatra and Seyranian’s (2016) framework, we found evidence of a knowledge-attitude link similar to prior research (Ecker et al., 2014; Heddy et al., 2016; Thacker et al., 2020). Specifically, readers with high levels of prior knowledge and positive attitudes (*positive-accurate*) were less likely to backfire compared to other groups. Building from this knowledge-attitude link, we documented the role of two epistemically related emotions. Anger was the only significant predictor of knowledge backfire and confusion was the sole significant positive predictor of attitudinal backfire. While the refutation texts themselves did not differentially elicit anger and confusion, significant differences emerged when we analyzed the emotions of those who demonstrated backfire or knowledge revision. While anger (M = 1.68) and confusion (M = 1.85) were experienced at low overall intensities (1 = *not at all*, 2 = *a little*), those who demonstrated knowledge backfire were angrier (M = 2.16; d = 0.49) and more confused (M = 2.19; d = 0.37) than those who revised their knowledge (M = 1.56 and M = 1.76 respectively). In line with D'Mello and Graesser (2014) and their view on confusion-frustration and confusion-boredom links and Immordino-Yang and Damasio's view of the emotional rudder, our results indicate that small differences in emotions experienced and their impact on learning processes are worthy of future consideration. It seems likely that anger that arose during the reading of refutation text could have contributed to preventing participants from revising their misconception, thereby leading to the detected backfire effects. Also, participants who were confused by the texts may have relied more heavily on their existing attitudes as compared with individuals who may not have found the text confusing. Despite the importance of these emotions in predicting backfire we failed to provide evidence that specific emotions arose from the supporting information provided.

Given our interest in understanding who demonstrated backfire and under what conditions, we examined those who revised or demonstrated backfired in knowledge and attitude separately. For those who revised their knowledge we documented relationships between prior knowledge, attitudes, and emotions seen in prior research on refutation texts. That is, prior knowledge was predictive of attitudes with the epistemic emotion surprise mediating relations between pre-attitudes and post-knowledge and both surprise and hope mediating relations between pre-attitudes and post-attitudes (Heddy et al., 2016; Thacker et al., 2020; Vogl et al., 2019). However, for those who demonstrated knowledge backfire, this relationship was not observed (in part due to limited sample size given few readers demonstrated a backfire effect). Although tentative, this finding supports Sinatra and Seyranian's (2016) framework for the relationship between knowledge and attitudes and contextualizes Immordino-Yang and Damasio (2007) statement "I feel therefore I learn." In participants who demonstrated knowledge backfire no emotions mediated the relationship, although confusion, anger, frustration, and anxiety were involved. In line with D'Mello and Graesser (2014), confusion is a common and recurrent emotion, one that has two outcomes, depending on whether the confusion can be resolved or not. If the confusion can be resolved, learning can occur, however if the confusion cannot be resolved, it may lead to frustration and boredom. This interpretation relies on future studies to investigate the self-regulatory processes readers use to resolve confusion while also attending to backfire effects. Additionally, anger's role in these revision and backfire processes may benefit from a plausibility judgment analysis, for Lombardi and Sinatra (2013) found that as judgements of implausibility increased so did anger. This anger may also stem from participants active resistance to knowledge or attitude revision as the topic may be too close to their identity (Trevors et al., 2016). Negative emotions may arise from discomfort with being confronted, dispositional differences in processing, or an awareness that knowledge and attitudes are misaligned. For whatever reason, the pathway to knowledge revision documented in prior research, specifically the positive relationship between knowledge and attitudes, did not emerge, nor did their mediation via epistemic emotions.

Implications and Future Directions

Given the danger of backfire effects in science communication, and evidence of their prevalence, concerns about refutation texts and backfire effects warrant further investigation. Our study is aligned with recent investigations that show either no or modest backfire effects (Lewandowsky et al., 2020; Schmid & Betsch, 2019; Wood & Porter, 2019). If only a small percentage of readers of a refutation text backfire, perhaps the threat of backfire should be considered modest, especially since the backfire effect may not be associated with the text structure itself. Refutation texts are documented to be effective in revising inaccurate information in formal, non-formal, and informal learning environments and may yet play a role in combatting misinformation online. Rather than recommend against their use, we suggest that additional ways to attenuate backfire effects should be explored to strengthen the text while guarding against potential backfire effects.

Refutation texts augmented with graphics (Danielson et al., 2016), persuasive information (Thacker et al., 2020), and emotional content (Trevors & Kendeou, 2019) have generally been found to increase their effectiveness. In addition to augmenting the text, enhancing the breadth and depth of the supporting information could be powerful. The amount of supporting information provided in the text used here was modest, and perhaps the power or danger of this information could be more systematically manipulated, while maintaining focus on the emotional experience of readers. On balance, science communicators must weigh the positive effects of confronting misinformation online with the possible negative effect of inducing a backfire effect as suggested by Lewandowsky et al. (2020).

Fortunately, a limited number of participants demonstrated a backfire effect in our study. However, even a small number is dangerous if the topic is consequential. For example, a text designed to confront mask wearing misconceptions about COVID-19 will likely induce a backfire effect in some readers but facilitate revision among others. In our study, approximately one in five readers demonstrated a backfire effect when reading such texts. However small a proportion this may seem, even if a relatively small proportion of the population backfire when presented with scientific information and persist in choosing not to wear a mask as a result of their misunderstandings, their subsequent behaviors can put countless lives at risk and contribute to super-spreader events that are emblematic of uninformed behaviors. Thus, following best practices in how to carefully craft messages to maximize myth busting while minimizing backfire should be followed carefully especially when the topic is one where miscommunication comes with high stakes. (For best practices see, Lewandowsky et al.'s 2020 Debunking Handbook). Our results suggest that future research should continue to investigate who demonstrates backfire and under what conditions, while paying attention to the role of emotions.

Limitations

Like all studies, this study has limitations. First, to avoid complexity in the design of multiple texts on multiple topics with different types of supporting information, we chose to use a single text topic, GMFs. However, each topic is likely to have different effects on individual readers dependent on their familiarity with the topic, whether the topic is controversial, or tied to readers' identity. Second, the demographic data error specifically for political affiliation limits our interpretations. Third, to maximize comparability across conditions, the amount of supporting information provided was modest (2 to 5 sentences). The lack of differential impact of the supporting information conditions could be due to the chosen information being insufficient. Longer texts with more supporting information, or more surprising or hopeful information may show different results. Fourth, related to text length, this study explored knowledge revision and backfire effects in a short time period. Although emotions tend to be experienced in the moment and fade after a delay, future studies should assess the changes over multiple time periods to understand the delayed impacts of emotions on the backfire effect when reading refutation texts with different supporting information. Fifth, the study is limited due to its operationalization of the backfire effect. Wood and Porter (2019) state the backfire effect is elusive and surrounded by measurement ambiguity. Additional backfire measures, such as agreement scales may produce different findings that are less sensitive to regression to the mean or inattention. Finally, given the lack of text structure findings, conclusions about the impact of supporting information on backfire effects are not definitive from the present investigation, and should be interpreted with caution.

Conclusions

Educators, scientists, and science communicators may wish to confront their readers' inaccurate knowledge about a variety of topics, from vaccination and GMF safety to our role in climate change. Refutation texts have been shown to be effective in confronting misconceptions but concerns about the degree to which individuals might backfire, or double down on their prior belief, have raised concerns about the use of the refutation text format. Our research suggests that backfire effects from reading a refutation text are a modest effect. Individuals who were most successful in avoiding backfire and revising their inaccurate knowledge had a positive relationship between their knowledge and attitudes and leveraged their emotions in the process. However, individuals who demonstrated a backfire effect did not have a significant relationship between their knowledge and attitudes. Further, backfire in knowledge was predicted by anger, and attitudinal backfire was predicted by confusion. Given the importance of epistemic emotions in revision processes and our findings that different knowledge attitude links (e.g., knowledge predicting attitudes, no relationship) elicited different epistemic emotions is worthy of future study. Future research should continue to explore the efficacy of various methods of promoting scientific understanding and resistance to misinformation, including how to engage readers' emotions in support of accurate scientific concepts.

References

Arguel, A., Lockyer, L., Kennedy, G., Lodge, J. M., & Pachman, M. (2019). Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments, 27*(2), 200-210. https://doi.org/10.1080/10494820.2018.1457544

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223-241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Bohn-Gettler, C. M. (2019). Getting a Grip: The PET Framework for Studying How Reader Emotions Influence Comprehension. *Discourse Processes, 56*(5-6), 386-401. https://doi.org/10.1080/0163853x.2019.1611174

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604.

Broughton, S. H., Sinatra, G. M., & Nussbaum, E. M. (2011). "Pluto Has Been a Planet My Whole Life!" Emotions, Attitudes, and Conceptual Change in Elementary Students' Learning about Pluto's Reclassification. *Research in Science Education, 43*(2), 529-550. https://doi.org/10.1007/s11165-011-9274-x

Brun, G., & Kuenzle, D. (2008). Introduction: A New Role for Emotions in Epistemology. In G. Brun, U. Doguoglu, & D. Kuenzle (Eds.), *Epistemology and Emotions*. Ashgate.

Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracin, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science, 28*(11), 1531-1546. https://doi.org/10.1177/0956797617714579

Cook, J., & Lewandowsky, S. (2011). *The Debunking Handbook*. University of Queensland. http://sks.to/debunk

Cordova, J. R., Sinatra, G. M., Jones, S. H., Taasoobshirazi, G., & Lombardi, D. (2014). Confidence in prior knowledge, self-efficacy, interest and prior knowledge: Influences on conceptual change. *Contemporary Educational Psychology, 39*(2), 164-174.

D'Mello, S. K., & Graesser, A. C. (2014). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 289–310). Routledge/Taylor & Francis Group.

D'Mello, S. K., Lehman, B., Pekrun, R., & Graesser, A. C. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Danielson, R. W., Sinatra, G. M., & Kendeou, P. (2016). Augmenting the Refutation Text Effect with Analogies and Graphics. *Discourse Processes, 53*(5-6), 392-414. https://doi.org/10.1080/0163853x.2016.1166334

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Mem Cognit, 42*(2), 292-304. https://doi.org/10.3758/s13421-013-0358-x

Gaskell, G., Bauer, M. W., & Allum, N. (1999). Worlds apart? The reception of genetically modified foods in Europe and the US. *Science, 285*, 384-387. https://doi.org/10.1126/science.285.5426.384

Guzzetti, B. J. (2000). Learning Counter-Intuitive Science Concepts: What Have we Learned from over a Decade of Research? *Reading & Writing Quarterly, 16*, 89-98. https://doi.org/10.1080/105735600277971

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting Conceptual Change in Science: A Comparative Meta-Analysis of Instructional Interventions from Reading Education and Science Education. *Reading Research Quarterly, 28*(2), 116-159. https://doi.org/10.2307 / 747886

Heddy, B. C., Danielson, R. W., Sinatra, G. M., & Graham, J. (2016). Modifying Knowledge, Emotions, and Attitudes Regarding Genetically Modified Foods. *The Journal of Experimental Education, 85*(3), 513-533. https://doi.org/10.1080/00220973.2016.1260523

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education, 1*(1), 3-10. https://doi.org/10.1111/j.1751-228X.2007.00004.x

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports, 6*, 1-11. https://doi.org/10.1038/srep39589

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising, 46*(1), 141-155. https://doi.org/10.1080/00913367.2016.1269304

Kendeou, P., & O'Brien, E. J. (2014). The Knowledge Revision Components (KReC) Framwork: Processes and Mechanisms. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. The MIT Press.

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 854-865. https://doi.org/10.1037/a0029468

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory and Cognition, 35*(7), 1567-1577. https://doi.org/10.3758/bf03193491

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge Revision Processes in Refutation Texts. *Discourse Processes, 51*(5-6), 374-397. https://doi.org/10.1080/0163853x.2014.913961

Kessler, E. D., Braasch, J. L. G., & Kardash, C. M. (2019). Individual Differences in Revising (and Maintaining) Accurate and Inaccurate Beliefs About Childhood Vaccinations. *Discourse Processes, 56*(5-6), 415-428. https://doi.org/10.1080/0163853x.2019.1596709

Lem, S., Baert, K., Ceulemans, E., Onghena, P., Verschaffel, L., & Van Dooren, W. (2017). Refutational text and multiple external representations as a method to remediate the misinterpretation of box plots. *Educational Psychology, 37*(10), 1281-1300. https://doi.org/10.1080/01443410.2017.1333574

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., & Zaragoza, M. S. (2020). The Debunking Handbook 2020. Available at https://sks.to/db2020. https://doi.org/10.17910/b7.1182

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological science in the public interest, 13*(3), 106-131. https://doi.org/10.1177/1529100612451018

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. (2016). Plausibility Judgments in Conceptual Change and Epistemic Cognition. *Educational Psychologist, 51*(1), 35-56. https://doi.org/10.1080/00461520.2015.1113134

Lombardi, D., & Sinatra, G. M. (2013). Emotions about Teaching about Human-Induced Climate Change. *International Journal of Science Education, 35*(1), 167-191. https://doi.org/10.1080/09500693.2012.738372

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*(10), 464-470. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*(3), 232-257. https://doi.org/10.1016/j.cogpsych.2006.09.006

Lorenzoni, I., Leiserowitz, A., De Franca Doria, M., Poortinga, W., & Pidgeon, N. F. (2006). Cross-National Comparisons of Image Associations with “Global Warming” and “Climate Change” Among Laypeople in the United States of America and Great Britain1. *Journal of Risk Research, 9*(3), 265-281. https://doi.org/10.1080/13669870600613658

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology: Revue Canadienne De Psychologie Experimentale, 55*(1).

Muis, K. R., Pekrun, R., Sinatra, G. M., Azevedo, R., Trevors, G. J., Meier, E., & Heddy, B. C. (2015). The curious case of climate change: Testing a theoretical model of epistemic beliefs, epistemic emotions, and complex learning. *Learning and Instruction, 39*, 168-183. https://doi.org/10.1016/j.learninstruc.2015.06.003

Munnich, E. L., & Ranney, M. A. (2019). Learning From Surprise: Harnessing a Metacognitive Surprise Signal to Build and Adapt Belief Networks. *Topics in Cognitive Science, 11*(1), 164-177. https://doi.org/10.1111/tops.12397

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior, 32*(2), 303-330. https://doi.org/10.1007/s11109-010-9112-2

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 3-31). American Psychological Association. https://doi.org/10.1037/13274-001

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion, 31*(6), 1268-1276. https://doi.org/10.1080/02699931.2016.1204989

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond Cold Conceptual Change: The Role of Motivational Beliefs and Classroom Contextual Factors in the Process of Conceptual Change. *Review of Educational Research, 63*(2), 167-199. https://doi.org/10.3102/00346543063002167

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Ranney, M. A., Munnich, E. L., & Lamprey, L. N. (2016). Increased wisdom from the ashes of ignorance and surprise: Numerically-driven inferencing, global warming, and other exemplar realms. In *Psychology of Learning and Motivation* (Vol. 65, pp. 129-182). Academic Press. https://doi.org/10.1016/bs.plm.2016.03.005

Rapp, D. N., & Braasch, J. L. G. (2014). Accurate and inaccurate knowledge acquisition. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing Inaccurate Information*. MIT Press.

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know. Updating text representations during narrative comprehension. *Memory and Cognition, 35*(8), 2019-2032.

Rapp, D. N., & Kendeou, P. (2009). Noticing and Revising Discrepancies as Texts Unfold. *Discourse Processes, 46*(1), 1-24. https://doi.org/10.1080/01638530802629141

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1-36.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 41*(5), 513-536.

Schmid, P., & Betsch, C. (2019). Effective strategies for rebutting science denialism in public discussions. *Nature Human Behavior, 3*(9), 931-939. https://doi.org/10.1038/s41562-019-0632-4

Silvia, P. J. (2010). Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts, 4*(2), 75-80. https://doi.org/10.1037/a0017081

Sinatra, G. M. (2005). The "Warming Trend" in Conceptual Change Research: The Legacy of Paul R. Pintrich. *Educational Psychologist, 40*(2), 107-115. https://doi.org/10.1207/s15326985ep4002_5

Sinatra, G. M., & Broughton, S. H. (2011). Bridging Reading Comprehension and Conceptual Change in Science Education: The Promise of Refutation Text. *Reading Research Quarterly, 46*(4), 374-393. https://doi.org/10.1002/rrq.005

Sinatra, G. M., Broughton, S. H., & Lombardi, D. (2014). Emotions in science education. In *International handbook of emotions in education* (pp. 415-436).

Sinatra, G. M., & Hofer, B. K. (2016). Public Understanding of Science: Policy and Educational Implications. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 245-253. https://doi.org/10.1177/2372732216656870

Sinatra, G. M., Kienhues, D., & Hofer, B. K. (2014). Addressing Challenges to Public Understanding of Science: Epistemic Cognition, Motivated Reasoning, and Conceptual Change. *Educational Psychologist, 49*(2), 123-138. https://doi.org/10.1080/00461520.2014.916216

Sinatra, G. M., & Seyranian, V. (2016). Warm Change about Hot Topics. In L. Corno & E. Anderman (Eds.), *APA Handbook of educational psychology*. APA Publications.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6 ed.). Pearson.

Thacker, I., Sinatra, G. M., Muis, K. R., Danielson, R. W., Pekrun, R., Winne, P. H., & Chevrier, M. (2020). Using persuasive refutation texts to prompt attitudinal and conceptual change.

Journal of Educational Psychology, 112(6), 1085-1099. https://doi.org/http://dx.doi.org/10.1037/edu0000434

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-502.

Tippett, C. D. (2010). Refutation Text in Science Education: A Review of Two Decades of Research. *International Journal of Science and Mathematics Education, 8*(6), 951-970. https://doi.org/10.1007/s10763-010-9203-x

Trevors, G. J., & Kendeou, P. (2019). The Effects of Positive and Negative Emotions on Knowledge Revision.

Trevors, G. J., Kendeou, P., & Butterfuss, R. (2017). Emotion Processes in Knowledge Revision. *Discourse Processes, 54*(5-6), 406-426. https://doi.org/10.1080/0163853x.2017.1312201

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes, 53*(5-6), 339-370.

Uzogara, S. G. (2000). The impact of genetic modification of human foods in the 21st century: A review. *Biotechnology Advances, 18*, 179-206.

Vaughn, A. R., & Johnson, M. L. (2018). Communicating and enhancing teachers' attitudes and understanding of influenza using refutational text. *Vaccine, 36*(48), 7306-7315. https://doi.org/10.1016/j.vaccine.2018.10.040

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019). Surprise, Curiosity, and Confusion Promote Knowledge Exploration: Evidence for Robust Effects of Epistemic Emotions. *Front Psychol, 10*, 2474. https://doi.org/10.3389/fpsyg.2019.02474

Will, K. K., Masad, A., Vlach, H. A., & Kendeou, P. (2019). The effects of refutation texts on generating explanations. *Learning and Individual Differences, 69*, 108-115. https://doi.org/10.1016/j.lindif.2018.12.002

Wood, T., & Porter, E. (2019). The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence. *Political Behavior, 41*(1), 135-163. https://doi.org/10.1007/s11109-018-9443-y

Table 1
Descriptive statistics of central variables by backfire.

Measure	**Full Sample** (n=328)		**Backfire in Knowledge** (n=68)		**No Backfire in Knowledge** (n=260)		**Backfire in Attitude** (n=42)		**No Backfire in Attitude** (n=244)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Prior Knowledge	4.09	2.22	4.81	2.52	3.91	2.11	3.07	1.92	4.23	2.22
Final Knowledge	5.69	2.75	2.93	2.73	6.37	2.29	3.88	2.78	6.19	2.43
Prior Attitude	3.13	1.15	3.16	1.11	3.12	1.16	3.08	1.00	3.11	1.16
Final Attitude	3.35	1.15	3.19	1.06	3.39	1.17	2.56	0.96	3.49	1.13
Emotions										
Angry	1.68	1.09	2.16	1.45	1.56	0.95	2.05	1.15	1.60	1.06
Hopeful	2.68	1.19	2.86	1.32	2.64	1.16	2.64	1.01	2.65	1.18
Anxious	1.8	1.09	2.16	1.31	1.72	1.02	2.26	1.01	1.68	1.05
Bored	1.63	1.02	2.05	1.29	1.52	0.91	1.93	1.07	1.57	0.98
Curious	3.12	1.16	3.26	1.16	3.08	1.15	2.90	1.05	3.12	1.17
Happy	2.38	1.23	2.62	1.34	2.32	1.19	2.50	1.33	2.33	1.20
Fear	1.76	1.12	2.12	1.26	1.67	1.07	2.29	1.24	1.65	1.08
Confused	1.85	1.1	2.19	1.23	1.76	1.05	2.33	1.10	1.75	1.07
Interested	3.27	1.17	3.31	1.20	3.26	1.17	2.98	1.16	3.30	1.17
Surprised	2.53	1.32	2.57	1.48	2.52	1.28	2.57	0.99	2.50	1.35
Frustrated	1.69	1.1	2.09	1.33	1.60	1.01	2.14	1.18	1.61	1.06
Hopeless	1.55	1	2.00	1.34	1.44	0.87	2.17	1.25	1.43	0.89
Enjoyment	2.21	1.22	2.47	1.23	2.15	1.21	2.38	1.25	2.16	1.21

Table 2

Intercorrelations between variables used in path analyses.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Prior Knowledge	1																
2. Final Knowledge	.50***	1															
3. Prior Attitude	.37***	.20***	1														
4. Final Attitude	.31***	.31***	.83***	1													
5. Angry	.34***	.41***	.35***	.40***	1												
6. Hopeful	.12*	.02	.50***	.56***	.13*	1											
7. Anxious	.32***	.35***	.29***	.38***	.65***	.08	1										
8. Bored	.15*	.26***	.03	.08	.50***	.09	.43***	1									
9. Curious	.03	.02	.19**	.33***	.09	.47***	.02	.02	1								
10. Happy	.01	.13*	.43***	.44***	.03	.61***	.00	.20***	.43***	1							
11. Fearful	.27***	.36***	.32***	.41***	.68***	.05	.68***	.45***	.00	.04	1						
12. Confused	.32***	.31***	.24***	.25***	.51***	.07	.55***	.45***	.20***	.15*	.56***	1					
13. Interested	.09	.02	.31***	.38***	.19**	.52***	.04	.14*	.65***	.44***	.12*	.14*	1				
14. Surprised	.37***	.16**	.12*	.04	.20***	.24***	.20***	.19**	.40***	.29***	.14*	.37***	.34***	1			
15. Frustrated	.27***	.37***	.36***	.45***	.77***	.16**	.65***	.50***	.08	.01	.73***	.48***	.14*	.15**	1		
16. Hopeless	.28***	.42***	.21***	.29***	.69***	.04	.63***	.54***	.03	.16**	.64***	.55***	.07	.24***	.73***	1	
17. Enjoyment	.00	.13*	.39***	.39***	.05	.56***	.02	.19***	.39***	.82***	.09	.10	.45***	.28***	.04	.19**	1

Table 3
Logistic regression models predicting knowledge and attitude backfire using logodds as the coefficient.

	Backfire in Knowledge			Backfire in Attitude		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Causal	0.465	0.469	0.728~	0.205	0.274	0.171
Example	0.409	0.417	0.342	0.106	0.085	-0.077
Prior Attitude		0.604~	0.639~		0.688*	0.910*
Prior Knowledge		0.774***	1.052***		0.221	0.283
Prior Attitude*Prior Knowledge		-0.171*	-0.189*		-0.163~	-0.164~
Angry			0.563*			-0.483~
Hopeful			0.222			0.14
Anxious			0.048			0.182
Bored			0.202			-0.186
Curious			0.16			-0.203
Happy			0.06			-0.034
Fearful			-0.201			0.134
Confused			0.201			0.454*
Interested			-0.158			-0.415~
Surprised			-0.027			-0.201
Frustrated			-0.101			0.106
Hopeless			0.266			0.465~
Enjoyment			0.066			0.206
Observations	294	294	294	286	286	286
Akaike Inf. Crit.	296.194	287.767	278.084	244.389	235.136	237.567

Note:~$p < 0.1$ *$p < 0.05$;**$p < 0.01$;***$p < 0.001$

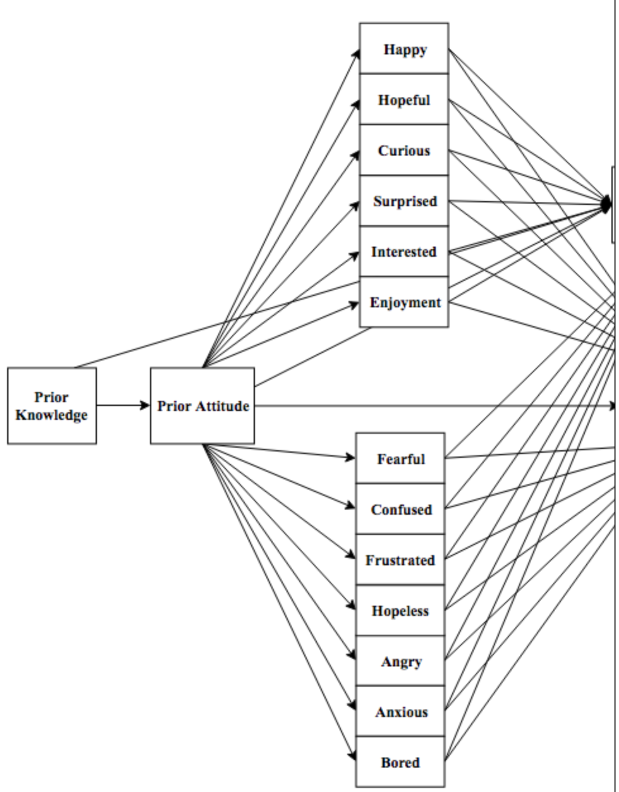

Figure 1. Hypothesized Model. Intercorrelations between emotions were included in the model but not shown here.

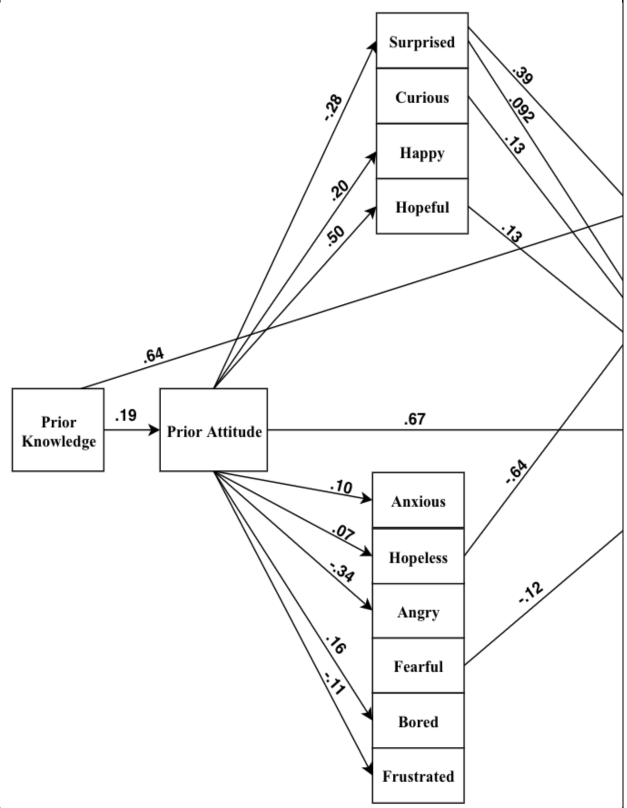

Figure 2. Model with full dataset. Only paths visible at the alpha < .05 level. Correlations between emotions not shown.

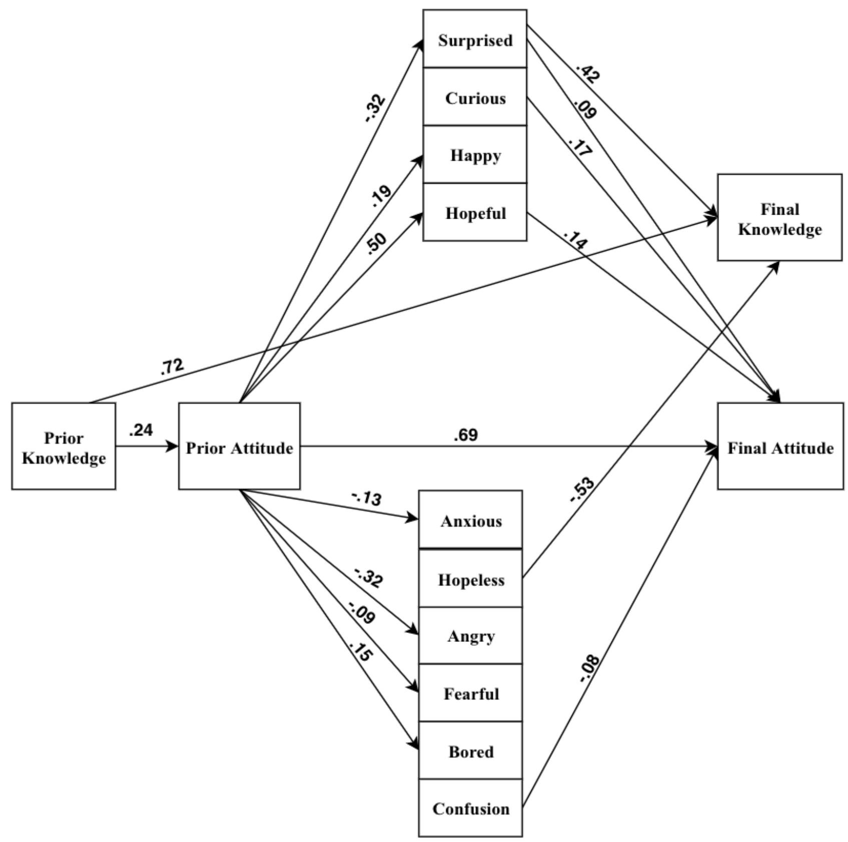

Figure 3.1 Model with readers who engaged in the **knowledge** revision process. Only significant relationships at the *alpha* < .05 level are shown. Correlations between emotions not shown.

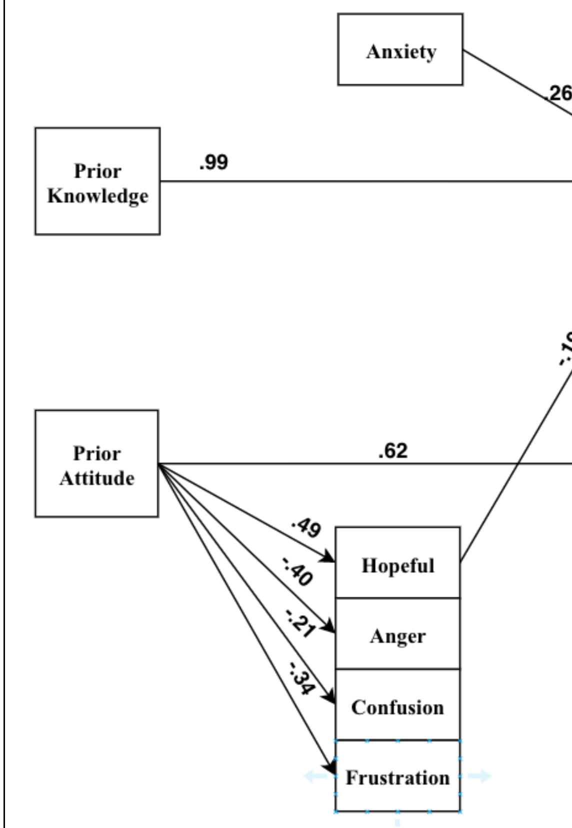

Figure 3.2 Model with readers who engage
processing in **knowledge**. Only significant
alpha < .05 level are shown. Correlations b
shown.

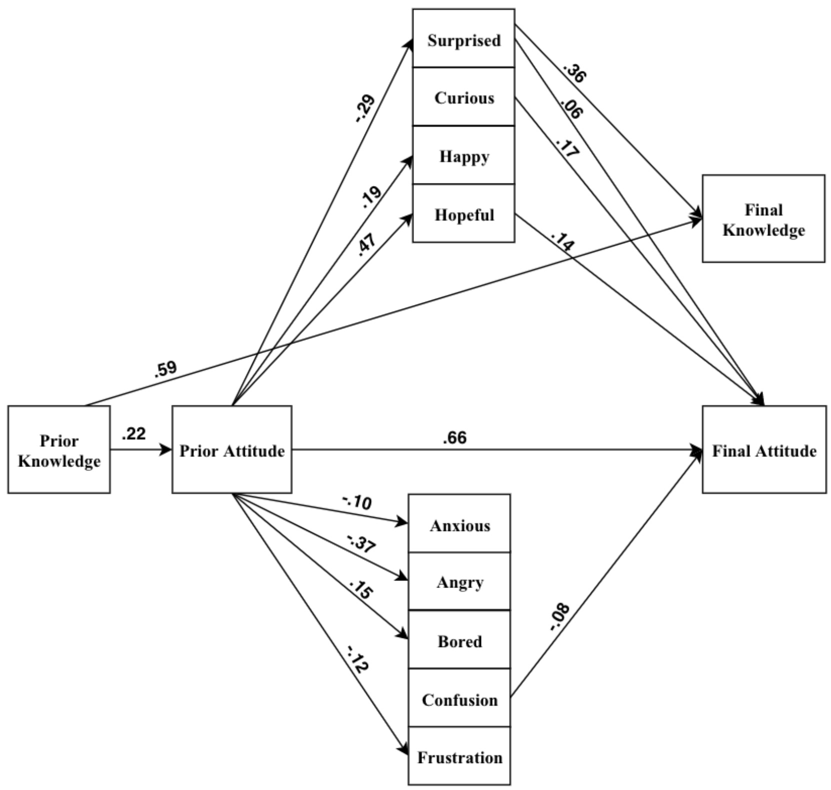

Figure 4. Model with readers who demonstrated a positive shift in **attitudes**. Only significant relationships at th
level are shown. Correlations between emotions not shown.