The Limits of Peer Influence: A Social (Dis)Affirmation Explanation of How Online Ratings Influence Trust in Factual Corrections

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

in the Graduate School of The Ohio State University

By

Rachel L. Neo, M.A.

Graduate Program in Communication

The Ohio State University

2016

Dissertation Committee:

R. Kelly Garrett, Advisor

David C. DeAndrea

David R. Ewoldsen

Erik C. Nisbet
Abstract

Political misperceptions pose a serious threat to democracy, making it imperative to understand how to correct such false beliefs. Online ratings could play an important role in this process. Research on bandwagon effects suggests that favorable online ratings should help make corrections more persuasive by fostering trust in such messages. The assumption that online ratings are uniformly persuasive is, however, overly simplistic. I argue that online ratings will not always promote acceptance of corrections. In what I term the social affirmation heuristic, I hypothesize that people will only trust ratings of factual corrections that affirm what they already believe, and vice versa. I further predict that rating trust will influence subsequent trust in corrections. Taken together, this means that belief discrepant ratings can have boomerang effects. Instead of eliciting the bandwagon effects described above, favorable ratings promote distrust of belief discrepant corrections. It is, however, possible for belief uncertainty and de-biasing messages to limit these boomerang effects by reducing reliance on the social affirmation heuristic. I expect these predictions to hold for various kinds of ratings, including both star ratings, which indicate rater favorability toward content, and Likes, which indicate the number of raters who see value in the content.

This dissertation uses two studies to test these ideas. The first study uses data from an online experiment conducted with convenience sample of 847 participants. The data for the second study come from a nationally representative sample of 500 participants. With the exception of the hypothesis that belief confidence affects rating
trust, all hypotheses received robust support in the context of star ratings. Implications of these findings are discussed.
Dedication

This work is dedicated to all the academics out there who keep trying, failing, and getting back up again. May you triumph against the odds and prove your detractors wrong.
Acknowledgements

Some of this material is based on work funded by the U.S. National Science Foundation under Grant No. IIS-1149599 and the Omnibus Survey for the School of Communication (OSoC). Any opinions, findings, and conclusions reflected in the material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF) or OSoC.

I would like to thank my advisor Dr. Kelly Garrett for pushing me to hone my writing and theorizing, and for expressing confidence in my scholarly abilities during moments of self-condemnation. I will always remember your wise words of wisdom, that a) things might not pan out in the short-run, but everything will work out in the long run, and b) to put the pursuit of knowledge above all else in academia. I aspire to make theoretical breakthroughs just like him someday.

I would also like to thank my dissertation committee members, Dr. Dave Ewoldsen, Dr. Erik Nisbet, and Dr. Dave DeAndrea for their constructive feedback and suggestion. This dissertation would not have come to fruition without their invaluable input.

I also wish to thank Dr. Jesse Fox for being my sounding board and giving practical advice on how to overcome imposter syndrome in academia. Moving forward, I hope to be a scholar like her who knows how to listen and build people up.

I am also very grateful to the faculty, staff, and graduate students of the Ohio State University’s School of Communication for their support and understanding throughout my five years as a doctoral student there. Thank you for helping me to grow as a scholar and a teacher.

Many thanks to my good friend and cohort buddy, Kelly Dillon, for your pep talks, delicious home-cooked meals, and Thanksgiving dinners. Thank you for helping me to adapt to life in Columbus, calling me out whenever I start to spiral, and being the very friend that I needed.

I would also like to give a big shout out to Dr. Mark Cenite for lending a non-judgmental listening ear to my rants from the other side of the Pacific. Thank you for being my mentor even after all these years and appreciating my sense of humor.

I am also very thankful for friends such as Brian and Sally Sansbury. Thank you for inviting me over to your place for hotpot during the (ironically) snowy Columbus winters of the Lunar New Year and praying for me in times of need.
My academic journey would not have come to fruition without the parasocial support from my favorite Disney animated features. These movies provided me with much comfort and solace whenever I faced academic struggles. Many thanks to *Frozen*, which taught me and still teaches me to ‘test the limits and break through’ and to ‘let go of fear’. Special credit goes to *Zootopia* for being a very relatable movie that reflects my struggles and triumphs as a foreign student in America. Someday, I hope to be just like Judy Hopps, and prove my naysayers wrong. And I will return to that city of opportunity and make it big there.

Most of all, I am extremely grateful to my mom and dad for their unflagging love, prayers, emotional support, and awesome local food. I love you both, and hope to do you proud someday.

Last but not least, I am thankful to God for seeing me through the ups and downs of these five years. Thank You for your provisions during trying times.
Vita

2007 ........................................ Bachelor of Communication Studies, Wee Kim Wee School of Communication & Information, Singapore

2010 ........................................ Masters of Communication Studies (By Research), Wee Kim Wee School of Communication & Information, Singapore

2011 ........................................ Graduate Teaching Associate, School of Communication, The Ohio State University
to present

Publications


Fields of Study

Major Field: Communication
Table of Contents

Abstract.................................................................................................................. ii
Dedication................................................................................................................ iv
Acknowledgements.................................................................................................. v
Vita.......................................................................................................................... vii
List of tables............................................................................................................ x
List of figures........................................................................................................... xi
Chapter 1: Introduction............................................................................................ 1
Chapter 2: Literature Review.................................................................................... 5
Chapter 3: Methods.................................................................................................. 16
Chapter 4: Results.................................................................................................... 40
Chapter 5: Discussion.............................................................................................. 61
References................................................................................................................ 72
Appendix A: Study 1’s De-biasing & Bias Inducing Messages............................... 80
Appendix B: Study 1 & Study 2’s Fact Checking Messages and Online Ratings 85
Appendix C: Stimulus Pre-Test Questions............................................................... 92
Appendix D: Study 1’s Survey Questions................................................................. 96
Appendix E: GFK Sample Description................................................................. 110
Appendix F: Study 2’s Survey Questions............................................................... 113
List of Tables

Table 1: Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Ratings on Perceived Similarity with Raters’ Beliefs Depend on Belief Consistency in Study 1 ................................................................. 44

Table 2: Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Ratings on Perceived Similarity with Raters’ Beliefs Depend on Belief Consistency in Study 2 ................................................................. 46

Table 3: Unstandardized Betas and Standard Errors For Perceived Similarity With Raters’ Beliefs As a Predictor of Online Rating Trustworthiness in Study 1. 49

Table 4: Hierarchical Linear Regression Analyses With Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Rating Favorability on Trust in Fact Checking Messages Depend on Rating Trust in Study 1 ................................................................. 51

Table 5: Unstandardized Betas and Standard Errors Examining Whether the Effect of Perceived Similarity With Raters’ Beliefs on Rating Trust Depends on Belief Confidence and Heuristic Reliability in Study 1 ......................... 55

Table 6: Conditional Direct Effects of Perceived Similarity With Raters’ Beliefs On Trust in Online Ratings at Various Values of Heuristic Reliability in Study 1. 58
List of Figures

Figure 1: Overarching Theoretical Model .......................................................... 10

Figure 2: Favorable Star Rating With Fact Checking Message On The MMR Vaccine And Autism .......................................................... 20

Figure 3: Conditional Direct Effect Of Star Rating On Perceived Similarity With Raters’ Issue Beliefs Among Participants Exposed To Belief Consistent Fact Checking Messages And Belief Inconsistent Fact Checking Messages In Study 1. 42

Figure 4: Conditional Direct Effect Of Star Rating On Perceived Similarity With Raters’ Issue Beliefs Among Participants Exposed To Belief Consistent Fact Checking Messages And Belief Inconsistent Fact Checking Messages In Study 2. 47

Figure 5: Trust in Star Ratings As A Significant Overall Moderator Of The Relationship Between Favorability Of Star Ratings And Trust In Fact Checking Messages. .......................................................... 53

Figure 6: Heuristic Reliability As A Significant Overall Moderator Of The Relationship Between Perceived Similarity With Raters’ Beliefs And Trust In Star Rating .......................................................... 57
Chapter 1: Introduction

Political misperceptions can potentially impede democracy. They prevent individuals from making informed decisions and stymie democratic discourse between opposing groups of individuals (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Nyhan & Reifler, 2010). Fact checking messages have been cited as a solution to correcting such misperceptions (Lewandowsky et al., 2012). However, the persuasive effects of these fact checking messages are limited, and are only likely to resonate among individuals who believe these messages (Lewandowsky et al., 2012).

Internet-based communication technologies such as social media platforms might contribute to the proliferation of misperceptions, but more importantly, they might lend some credence to corrections (Garrett & Weeks, 2013). It is thus imperative to examine how social media cues enhance the persuasiveness of fact checking messages. In particular, this study focuses on examining how online ratings, a prominent type of consensus cue on social media platforms, increase trust in fact checking messages.

Online ratings are explicitly intended to promote consensus about objects, people, or issues (Walther & Jang, 2012; Van Der Heide & Lim, 2015). Research has largely documented the powerful and persuasive effects of these ratings (e.g., Messing & Westwood, 2012; Sundar, Xu, & Oeldorf-Hirsch, 2009). In what has been termed as the “bandwagon heuristic”, most research on online ratings has assumed that peoples’ reactions toward online content will mirror rating scores (e.g., Sundar &

However, the aforementioned studies tend to focus on examining the effects of ratings on decisions to invest in goods or services that people have little prior beliefs about. Little research has examined how online ratings influence perceptions of message content that people might have strong existing beliefs about, such as ‘Likes’ given to a news story on Donald Trump or star ratings about a book about Obama’s leadership skills. Online ratings that accompany goods and services might work differently from ratings that accompany such message content. Ratings accompanying goods and service typically help individuals to decide whether to consume some product or service (Sundar et al., 2009). Individuals are unlikely to have strong beliefs about goods and services that they have not consumed. By contrast, individuals are likely to have existing beliefs or attitudes about messages that they have already consumed, especially fact checking messages aimed at countering erroneous beliefs on controversial socio-political issues. Consequently, individuals are more likely to have unwavering trust in online ratings accompanying goods and services than online ratings accompanying politically divisive fact checking messages.

This dissertation focuses on ratings that accompany fact checking messages, and asks whether the ratings influence the consumer’s response to these fact checking messages.

Individuals tend to hold very strong beliefs on controversial socio-political issues (Lewandowsky et al., 2012). They are often motivated to defend these issue
beliefs (McGuire & Papageorgis, 1962; Taber & Lodge, 2006). They will first gauge the degree to which others share these beliefs (Fields & Schuman, 1976). Individuals are more likely to feel that their beliefs are threatened when others hold dissimilar beliefs. To defend their beliefs, they will distrust belief inconsistent fact checking messages and vice versa (Lewandowsky et al., 2012). These findings suggest that trust in online ratings of fact checking messages depends on the extent to which these ratings come from individuals who affirm consumers’ issue beliefs.

In what I term as the “social affirmation” heuristic, I argue that individuals will first use their issue beliefs and online ratings in tandem to assess whether these ratings come from like-minded individuals. The more raters are perceived to hold dissimilar issue beliefs, the more an individual will feel that his or her beliefs are threatened. They will then defend threatened issue beliefs by distrusting belief discrepant online ratings, and protect their beliefs by trusting belief affirming online ratings. In addition, trust in online ratings is only as important as its influence on the persuasiveness of the message being rated. I further examine whether rating trust influences evaluations of fact checking message trustworthiness.

Also, I examine whether belief confidence and perceived heuristic reliability can influence individuals’ reliance on the social affirmation heuristic. Individuals who are very confident about their beliefs are likely to distrust information that challenges these beliefs, and to also readily trust information that affirms these beliefs than those who lack belief confidence (Koehler, 1993). This suggests that belief confidence will influence the degree to which individuals will rely on the social affirmation heuristic.
when evaluating rating trustworthiness. Also, research indicates that heuristic
reliability, i.e., the extent to which a heuristic is deemed as a valid guide for decision
making, can moderate the heuristics’ influence (Chen & Chaiken, 1999). This
suggests that the perceived reliability of the social affirmation heuristic will also
influence the extent to which individuals rely on this heuristic when evaluating the
trustworthiness of online ratings.

I use two studies to test out these aforementioned hypotheses. The first study
uses data from an online experiment conducted with a convenience sample of 847
participants from Qualtrics Panels. This study examines whether there is evidence for
the social affirmation heuristic, gauges whether rating trust influences message trust,
and tests whether belief confidence and heuristic reliability can influence reliance on
the social affirmation heuristic. The data for the second study come from a nationally
representative sample of 500 participants from GFK. This study gauges whether the
general population relies on the social affirmation heuristic when gauging belief
similarity with raters.
Chapter Two: Literature Review

Fact Checking Messages

Individuals often hold erroneous beliefs about factual evidence on controversial issues (Lewandowsky et al., 2012). Such erroneous beliefs hinder sound political decision making (Garrett & Weeks, 2013; Lewandowsky et al., 2012; Nyhan & Reifler, 2010). Consequently scholars and political practitioners have attempted to use fact checking messages to combat these misperceptions (Lewandowsky et al., 2012). These fact checking messages can either appear as standalone entities on fact-checking websites such as Factcheck.org or form part of the narrative in a newspaper article or political advertisement.

Research has shown that fact checking messages have limited persuasive efficacy because individuals often defend their beliefs by arguing against message claims (Lewandowsky et al., 2012). Scholars have suggested that consensus cues can either augment or undermine the persuasiveness of these fact checking messages (Lewandowsky et al., 2012). It is likely that belief defense motivations can either enhance or undermine trust in consensus cues such as online ratings of fact checking messages. In turn, trust in online ratings can affect the persuasiveness of fact checking messages. However, the mechanisms through which online ratings influence the persuasiveness of fact checking messages remain relatively unexplored. The next few sections of this chapter outline how ratings enhance or undermine the persuasiveness of fact checking messages.
Online Ratings

Online ratings are consensus cues that provide summative statistics of other users’ opinions or behaviors on issues, individuals, or objects (Walther & Jang, 2012). As with traditional types of consensus cues, online ratings ideally should help to forge consensus on messages, issues, or objects. Web-based collaborative filtering technologies allow summative statistics to be collated and displayed succinctly (Messing & Westwood, 2012; Walther & Jang, 2012; Van Der Heide & Lim, 2015). These online ratings might differ in terms of presentation format, e.g., counts versus text or proportions versus sheer number of views (Walther & Jang, 2012).

For purposes of this dissertation, it is useful to distinguish between two main types of online ratings. As outlined above, there are online ratings that help individuals to decide what types of content to consume, such as favorable online ratings that persuade individuals to choose to watch a movie that they know little about, online ratings that encourage individuals to click on the headline of a news story and read it, or online ratings indicating that a product has received favorable evaluations help individuals to decide whether to buy it (e.g., Sundar & Nass, 2001; Sundar et al., 2007; Messing & Westwood, 2012). Second, there are online ratings that are ostensibly aimed at helping individuals to decide whether content that has already been consumed can be trusted (Walther, Liang, Ganster, Wohn, & Emington, 2012). For instance, favorable ratings of a news story are aimed at persuading individuals to trust this story. In this study’s context, online ratings accompanying
full-length fact checking messages are supposed to help promote acceptance of these fact checking messages.

Research has shown that online ratings can have powerful persuasive effects (e.g., Sundar et al., 2009; Sundar et al., 2007; Messing & Westwood, 2012). However, most studies have been conducted in contexts involving decisions to select or consume content (e.g., Messing & Westwood, 2012; Van Der Heide & Lim 2015). Little research has examined how online ratings influence trust in message content that has already been consumed. In these instances, people are likely to already have existing beliefs about such message content. This is especially so in the context of messages aimed at countering deep-seated misperceptions about controversial socio-political issues, e.g., star-ratings of a fact checking message. Under these circumstances, online ratings might fail to forge consensus about the message’s trustworthiness. In the following sections of this chapter, I explain how individuals’ existing beliefs influence rating trust. I also examine how trust in online ratings in turn affects trust in the fact checking message, and how rating trust can be increased.

**The Social Affirmation Heuristic**

Current research tends to assume that individuals uniformly trust online ratings that inform content consumption (Messing & Westwood, 2012; Sundar & Nass, 2001). However, in what I term the social affirmation heuristic, I posit that individuals’ trust in online ratings of a factual claim will be shaped by their prior beliefs on the topic. This heuristic unfolds in two steps: Individuals will first use rating scores to gauge belief similarity with raters. Next, they will question the
trustworthiness of online ratings from raters with dissimilar beliefs, and will only trust online ratings from raters affirming their beliefs.

Individuals are compelled to protect their issue beliefs (Taber & Lodge, 2006). When beliefs are threatened by belief inconsistent information, it is likely that individuals will take steps to defend these beliefs by rejecting such information. And when beliefs are affirmed by belief consistent information, individuals will likewise take steps to preserve these beliefs by accepting such information (Taber & Lodge, 2006; Lewandowsky et al., 2012). Online platforms do not eliminate the occurrence of such biased processing. A series of exploratory focus groups indicates that individuals are likely to only trust information from online sources who affirm their beliefs (Metzger, Flanagin, & Medders, 2010). With this in mind, I coin the term ‘social affirmation heuristic’ to explain how individuals defend their beliefs in the face of a fact-checking message. The heuristic entails two key functions. As outlined in Figure 1 below, individuals will first use rating scores to determine whether raters share their beliefs. Second, because individuals are motivated to defend their beliefs, they will next use their perceptions of belief similarity with raters to decide whether such cues are trustworthy.

To determine the degree to which online ratings threaten one’s beliefs, individuals will first assess belief similarity with raters. Individuals are inherently ‘social’ creatures who desire to fit in (Asch, 1951; Gunther & Storey, 2003; Noelle-Neumann, 1974). They desire to know if their issue beliefs are widely shared, and
regularly assess the extent to which others share their issue beliefs (Fields & Schuman, 1976).

Online ratings are consensus cues that provide summative statistics indicating the degree to which others have similar beliefs. It is likely that consumers will infer belief similarities with raters from the online rating score. Furthermore, the balance theory can explain how people conclude that raters share their beliefs. According to the balance theory, people are motivated to achieve cognitive consistency when making evaluations (Cartwright & Harary, 1956). They will make interrelated favorable (+) or unfavorable (-) evaluations about other people and objects in order to maintain such cognitive consistency (Cartwright & Harary, 1956). Cognitive consistency is attained when the product of these evaluations is positive (Cartwright & Harary, 1956). For example, a person who likes (+) an object will conclude that others who evaluate that object favorably (+) are likeable (+). Multiplying these three positive signs together yields a consistent cognitive state. Conversely, a person who likes (+) an object will conclude that those giving unfavorable (-) object evaluations are dislikable (-). Likewise, multiplying these valenced evaluations together indicates cognitive balance.

Extrapolating this to the context of online ratings of fact checking messages, people might strive to maintain cognitive consistency when deciding if raters share their beliefs. They are likely to conclude that raters who give favorable (+) ratings to belief discrepant (-) fact checking messages have dissimilar beliefs (-). Similarity, unfavorable (-) ratings of belief consistent (+) fact checking messages might signal
belief dissimilarity with raters (-). By contrast, favorable (+) ratings of belief consistent messages (+) are likely to indicate belief similarity with raters (+). Unfavorable (-) ratings of belief inconsistent (-) messages might also cause individuals to draw similar conclusions (+). I hypothesize that:

H1a: Perceived similarity with raters of belief consistent messages will increase as rating scores increase.

H1b: Perceived similarity with raters of belief inconsistent messages will decrease as rating favorability increases.

---

**Figure 1.** Overarching theoretical model.

The greater the level of perceived belief dissimilarity with others, the more belief threat the consumer experiences. Individuals are assuaged when their opinions are in line with what most others think (Noelle-Neumann, 1974). The more others are
perceived to disagree with one’s opinions, the more individuals will feel that their opinions are threatened (Noelle-Neumann, 1974). An individual’s opinions and attitudes are often indicative of their issue beliefs (Eagly & Chaiken, 1998). When their opinions are threatened, it is likely that individuals will also feel that their beliefs are under attack. The more others hold dissimilar beliefs, the more they will feel that their beliefs are threatened. In this study’s context, the more individuals perceive that raters have dissimilar beliefs, the more they will feel that their beliefs are threatened. Conversely, the more individuals perceive that raters share their beliefs, the less they will feel that their beliefs are threatened.

Next, individuals will attempt to cope with perceived threats to their issue beliefs by judging rating trustworthiness. Questioning information trustworthiness is a common way of defending threatened issue beliefs. In contrast, accepting information trustworthiness is a common way of preserving issue beliefs that are not threatened (Lewandowsky et al., 2012). In what scholars have dubbed as the “belief force equals credible source” heuristic, individuals will attribute belief inconsistent information to untrustworthy sources and belief consistent information to trustworthy sources (Fragale & Heath, 2004). From this, it is likely that individuals will question rating trustworthiness in order to defuse perceived threats to their beliefs. If individuals perceive belief dissimilarity with raters, they will dismiss these ratings as being untrustworthy. Conversely, individuals will trust of online ratings in order to preserve their existing beliefs. If individuals perceive belief similarity with raters, they will deem these ratings to be trustworthy.
H2: Rating trust increases as perceived belief similarity with raters increases.

**The Moderating Role of Online Rating Trust on Trust in Fact Checking Messages**

The perceived trustworthiness of online ratings is only as important as its influence on the trustworthiness of the message being rated. As illustrated in Figure 1 above, the effects of rating favorability on perceived trust in fact checking messages depends on rating trust. The more an individual trusts online ratings, the more likely they will use them when evaluating the trustworthiness of fact checking messages. Conversely, the less individuals trust online ratings, the more likely these ratings will elicit boomerang effects. When this happens, individuals will form evaluations of fact checking message trustworthiness that conflict with rating scores.

When individuals trust a source, they are likely to adopt issue attitudes that are in line the source’s issue attitudes. Conversely, individuals are less likely to rely on untrustworthy sources when forming attitudes on issues (for a review see Pornpitakpan, 2004). Research has shown that individuals will develop positive product attitudes when they read favorable product reviews from trusted reviewers, and vice versa (Van Der Heide & Lim, 2015). This suggests that individuals who trust ratings are likely to form message evaluations that concur with the overall rating score. Favorable online ratings will lead to trust in fact checking messages whereas unfavorable online ratings will cause message distrust. However, this positive relationship between online ratings and trust in fact checking messages is likely to be weaker among individuals who distrust online ratings.
H3a: The less individuals trust online ratings, the less positive the association between online ratings and trust in fact checking messages.

In addition, it is possible for low levels of trust in online ratings to cause these online ratings to elicit boomerang effects. The social affirmation heuristic suggests that individuals will distrust online ratings that disaffirm issue beliefs. When this happens, favorable ratings can produce boomerang effects by reducing trust in fact checking messages instead of eliciting trust. The inverse is also true: Unfavorable ratings can ironically promote message trust.

The following hypothesis is also proposed:

H3b: Low levels of trust in online ratings will result in a boomerang effect, creating a negative association between online ratings and trust in fact checking messages.

Increasing Trust in Online Ratings

As outlined in the section above, distrust in online ratings will elicit boomerang effects when individuals evaluate the trustworthiness of fact checking messages. Instead of helping individuals to form conclusions about message trustworthy, distrusted online ratings are likely to be viewed with suspicion. This in turn raises the question of how to attenuate these effects, and increase trust in online ratings of fact checking messages. As outlined in Figure 1 above, I examine how two factors, belief confidence and perceived heuristic reliability, influence reliance on the social affirmation heuristic when individuals evaluate rating trustworthiness.
The influence of belief confidence on trust in online ratings. Individuals who hold very confident beliefs on controversial issues are also uniquely prone to being skeptical about media content. They are likely to question the trustworthiness of online ratings that disaffirm their beliefs and trust ratings that affirm their beliefs.

In general, individuals who are confident of their beliefs are certain that these beliefs are accurate (Petty, Brinol, & Tormala, 2002). These individuals are also more likely to hold strong issue beliefs (Marks & Kamins, 1988). Research has shown that individuals with strong issue beliefs are more likely to dismiss belief inconsistent scientific evidence and embrace belief consistent scientific evidence than those with weak issue beliefs (Koehler, 1993).

In light of these findings, it is unsurprising that the stronger an individual’s issue beliefs are, the more likely he or she will distrust belief inconsistent media content. Since belief strength and belief confidence are generally correlated (Marks & Kamins, 1988), belief confidence should also promote skepticism toward belief inconsistent media content and trust towards belief consistent media content. This implies that belief confident individuals are more likely to distrust belief discrepant ratings than individuals who lack belief confidence. They will also be more likely to trust ratings from raters who affirm their beliefs than belief unconfident individuals. I hypothesize that:

H4: The more confident an individual’s beliefs, the more positive the relationship between perceived similarity with the rater and online rating trust.
The influence of perceived heuristic reliability on trust in online ratings.

The perceived reliability of a heuristic refers to the extent to which a person trusts a heuristic to be a valid guide in making judgments (Chen & Chaiken, 1999). If individuals perceive a heuristic to be high in reliability, they will be more likely to use it in belief relevant situations. Conversely, when individuals do not deem a heuristic to be a reliable decision rule, they will be less likely to apply that heuristic in a given context (Kunda & Nisbett, 1986; Tversky & Kahneman, 1971).

Although individuals are prone to rely on cognitive biases when evaluating politically divisive fact checking messages, research has shown that de-biasing messages can reduce dependence on such biases (Schul, 1993). In this context, these de-biasing messages make individuals aware that relying on the social affirmation heuristic can lead to inaccurate conclusions about rating trustworthiness. Individuals who are exposed to a de-biasing message might be less likely to distrust belief discrepant online ratings, and less likely to trust belief consistent online ratings. By contrast, these findings also suggest that bias inducing messages can increase the likelihood of heuristic reliance. Such messages might further encourage individuals to distrust belief discrepant ratings and trust belief consistent ratings.

I hypothesize that:

H5: As heuristic reliance increases, the more positive the relationship between perceived similarity with the rater and online rating trust.
Chapter 3: Methods

Overview

I use two experiments to test the hypotheses outlined in the previous chapter. The first is an online experiment with a diverse convenience sample. It asks whether perceived similarity with raters’ beliefs influences rating trust. This experiment also examines whether the relationship between rating favorability and trust in a fact-checking message depends on rating trust. Finally, this study examines whether belief confidence and reduced reliance on the issue belief-rating trust heuristic can increase rating trust. The second is a survey-experiment conducted with a representative national sample. The experiment examines whether the general population infers raters’ beliefs from the ratings they assign to belief-relevant content.

Study 1: Online Experiment

Design. Study 1 utilized a 2 (Fact checking message: Belief consistent versus Belief inconsistent) x 2 (Star rating: Favorable vs unfavorable star rating) x 2 (Likes: Many vs. few Likes) x 2 (Heuristic reliability high versus heuristic reliability low) between-subjects experimental design. Participants were randomly assigned to conditions.

Sample. There were 847 participants in Study 1. These participants were part of an opt-in panel managed by Qualtrics. Panelists agreed to participate in studies in exchange for incentives. In order to approximate a representative sample, participants were recruited such that the sample is matched on key Census demographics such as age, $M = 49.1$, $SD = 14.2$, gender (50% male), race (83.9% White), income, education
(80% obtained at least a high school diploma), and political background (45% were Democrats or Democrat leaning, 20% were Independent, and 25% were Republican or Republican leaning).

**Procedure.** Participants were told that they will be helping to evaluate the features of an online education portal, and that the evaluation study takes approximately 15-20 minutes of their time. Next, they answered questions about their beliefs and confidence of their beliefs on several issues, including three questions on the MMR vaccine-autism link, Obama’s vacation days, and nuclear power plant emissions, which are the focus of this study.

After that, participants were randomly assigned to view a page explaining how to interpret online ratings (in the form of star ratings or Likes). There are four versions of these instructions, one containing a message aimed at increasing trust in star ratings (Likes), stating that ratings are valid indicators of message trustworthiness because they often align with expert consensus. The other a message aimed at reducing trust in star ratings (Likes), stating that ratings are not valid indicators of message trustworthiness because they seldom align with expert consensus (exact message wording is available in Appendix A). Participants have to read all the information on the page before proceeding.

They were then randomly assigned to view an image of the (fictional) online educational portal containing one of three fact checking messages. The message was accompanied by one of five randomly assigned online ratings, which are described in
detail below (see Appendix B for screenshots of the fact checking messages with ratings).

Lastly, participants complete a questionnaire that includes manipulation check items, and items pertaining to perceived similarity between raters’ beliefs and issue beliefs, the trust of the star ratings, trust in the original message, arguments about the fact checking messages, and demographics. Participants were then thanked and debriefed.

Stimuli.

*Fact checking messages.* Participants were randomly assigned to one of three different fact checking messages: 1) Obama took fewer vacation days than George W. Bush, 2) The MMR vaccine does not cause autism, or 3) Nuclear power plants emit minimal levels of carbon dioxide. The full text of the fact checking messages can be found below.

These fact checking messages were roughly equal in terms of word count, and were adapted from fact checking websites such as FactCheck.org and Politifact. They were chosen because these topics have generated heated debates between opposing parties. Furthermore, these topics are representative of information targeting erroneous beliefs in three different content areas: contentious political, scientific, and health issues.

*Online rating conditions.* Online ratings come in various forms, such as star ratings or the number of Likes (Walther and Jang, 2012). Research has shown that individuals’ reactions depend on the type of web interface cue (Fu, 2012). Although
rating type is not the central focus of this study, participants might have different reactions depending on rating type. With this in mind, I used stimulus sampling, examining whether favorable star ratings and numerous Likes elicit similar effects, and whether the effects of unfavorable star ratings and relatively few Likes have comparable effects. With star-based ratings there are two possible scores: favorable or unfavorable star ratings. And with Likes, there are two possible scores: a ‘large’ number of Likes’, and a ‘small’ number of Likes.

There were five different rating conditions: In the favorable star rating condition, there was a yellow star rating of 4.5 out of 5 stars. By contrast, in the unfavorable star rating condition, there were yellow star ratings of 1.5 out of 5 stars. In addition, there was a condition with ‘many’ Likes in which a fact checking message receives 1059 Likes, a condition with a small number of Likes in which a fact checking message receives 61 Likes, and a condition with no online ratings which serves as a control. In all five online rating conditions, the number of participants who view the fact checking message was kept constant at 10,123. A sample fact checking message with favorable star ratings can be found in Figure 2 below. Images of the remaining ratings can be found in Appendix B.
The MMR Vaccine and Autism

10,123 people viewed the information below.

Question: Is the MMR vaccine likely to cause autism?

Answer: There is no scientifically proven link between the MMR vaccine and autism. Reports from both the American Academy of Pediatrics and the Centers for Disease Control and Prevention conclude that there is no proven association between the MMR vaccine and autism.

Autism is a chronic developmental disorder, often first identified in toddlers from age 18 months to 30 months. The MMR vaccine is administered just before the peak age of onset of autism symptoms. This timing leads some parents to mistakenly assume a causal relationship. There is no evidence that the MMR vaccine causes autism.

Figure 2. Favorable star rating with fact checking message on the MMR vaccine and autism.
**Heuristic reliability manipulation.** Before reading the fact checking message and their online ratings, participants are randomly assigned to view 1) A de-biasing message geared toward increasing trust in online ratings or 2) A bias inducing message geared toward reducing trust in online ratings. As described in the Procedures section above, the first message explains that online ratings are valid indicators of message trustworthiness because they tend to reflect expert consensus on issues, while the second explains that online ratings often diverge from experts’ beliefs (see Appendix A).

**Pretesting the stimuli.** Prior to administering Study 1, I pre-tested the fact checking messages (\(N = 60\)) and heuristic reliability manipulations (\(N = 40\)) on Mechanical Turk. This pre-test took approximately three minutes. Each of these participants was paid a cash incentive of $0.25. The results of these pretests follow.

**Fact checking message stimuli pre-test.** The fact checking message pre-test aimed to ensure that the three fact checking messages are comparably trustworthy, so participants were only shown the messages themselves, and not online ratings or the heuristic reliability manipulations. After seeing fact checking messages, participants answer five questions gauging message trust adapted from Sundar and Nass (2001). These items measure the extent to which participants perceive the message they just read to be accurate, believable, biased, fair, and objective, on five-point Likert scales, with 1 = Strongly Disagree to 5= Strongly Agree, and were summed up to form an index of message trust, \(M = 18.5, SD = 4.2, \text{Cronbach alpha} = .92\). A one-way
ANOVA shows that there are no statistically significant differences in perceptions of message trustworthiness across the three messages, \( F(2, 61) = 1.32, MSE = 0.70, p = .28, \eta^2 = 0.04. \)

Also, stimulus test participants answered questions such as “The information that you just read said that Obama had fewer (more) vacation days than Bush,” “The MMR vaccine is likely (unlikely) to cause autism,” and “Nuclear power plants emit a lot of (a little bit of) carbon dioxide gas,” on three-point scales featuring correct, incorrect, and “don’t know” options. These items were designed to assess whether participants can accurately recall what the message said. Across all three message conditions, the vast majority (approximately 90%) of participants were able to accurately recall the message content.

Again in Study 1, I checked whether participants in all three message conditions perceive these messages to be equally trustworthy. Unlike this stimulus pre-test, a one way ANOVA showed that the fact checking message on MMR vaccines is perceived to be significantly more trustworthy, \( M = 18.3, SD = 3.20, \) than the message on Obama’s vacation days, \( M = 17.5, SD = 3.58, \) or nuclear power plant emissions, \( M = 17.7, SD = 2.90, F(2, 844) = 4.91, MSE = 10.5, p < .01, \eta^2 = 0.12. \)

Thus, the hypotheses were tested once with all three issues and again with just the messages on nuclear power plant emissions and Obama’s vacation days. There are no substantively important differences in findings.

**Heuristic reliability manipulation pre-test.** I hypothesize that the effect of perceived similarity on rating trust should be stronger among those induced to rely on
the heuristic than among those induced to believe that the heuristic is unreliable. Before testing this hypothesis, it was necessary to pretest the stimuli to confirm that participants who are induced to rely on the heuristic will perceive ratings as significantly more unreliable indicators of message trust than participants who are induced to believe that heuristic is unreliable. First, participants viewed either the de-biasing or bias inducing message. Next, they answered question gauging whether the heuristic manipulation pre-test works. Four items measured whether this manipulation works, on five point scales anchored by 1= Strongly Disagree and 5 = Strongly Agree. They are: “Messages with negative online ratings are more likely to be untrustworthy,” “Messages with positive online ratings are more likely to be trustworthy,” “Online ratings are always valid indicators of message trustworthiness”, and “Online ratings are always reliable indicators of message trustworthiness”. These four items were summed to form an index, $M = 11.78$, $SD = 3.94$, Cronbach alpha = 0.86.

Also, I created a dichotomous independent variable in which participants in the ‘high heuristic reliability’ condition were coded as ‘0’, whereas those assigned to the ‘low heuristic reliability’ condition were coded as ‘1’. An independent samples $t$ test shows that the heuristic reliability manipulation works. Participants in the ‘high heuristic reliability’, i.e., bias inducing message condition, are significantly less likely to agree that online ratings are valid indicators of message trustworthiness, $M = 9.29$, $SD = 3.39$, than participants in the ‘low heuristic reliability’ i.e., de-biasing pre-test condition, $M = 14.28$, $SD = 2.67$, $t(40) = -5.31$, $p < .001$. 
A complete list of items in the stimulus pre-testing stage can be found in Appendix C.

**Data cleaning for Study 1.** This study used participants from an online panel which comprises self-selected participants. Consequently, such online panel data might contain measurement errors from participants who give cursory responses (Peifer & Garrett, 2014). Before running descriptives and scale reliability analyses on variables used for testing hypotheses, I applied four filters to filter out participants who did not pay attention to the experimental stimuli. Participants were excluded if they (1) took an unreasonable amount to time to complete the study, (2) failed to notice or (3) accurately recall the online ratings, and (4) could not identify their message condition accurately. More details about each of these criteria are provided below.

**Time filter.** First, participants who took suspiciously short or overly long periods of time to complete the study were screened out. It is likely that participants who take very little time to complete the study are not paying enough attention to the questions or experimental stimuli. Also, participants who took an unusually long time to complete the study were likely to have allowed their attention to wander or are distracted by other tasks that were not related to the survey. As such, only participants who took within 7-40 minutes to complete the experiment were included in the study. The 7\textsuperscript{th} minute timing ‘floor’ corresponded with the 10\textsuperscript{th} percentile of timings and this 40\textsuperscript{th} minute timing ‘ceiling’ corresponded approximately with the 95\textsuperscript{th} percentile of timings across the star ratings, Likes and no rating conditions.
**Message condition filter.** In addition to checking whether participants pay attention to the online ratings, I ensured that participants were paying attention to the *fact checking message* they were reading. Consequently, only participants who correctly identified the message condition to which they were assigned to were included in subsequent statistical analyses.

**Two online rating filters.** Participants were required to pay attention to the online ratings. The third filter question simply asked participants if they noticed the online rating on a dichotomous ‘Yes/No’ scale. The fourth filter criterion was more stringent and gauged whether participants can recall the numeric value of the online rating associated with the fact checking message they saw. Participants who did not notice the star rating or Likes or were unable to recall the numeric value of the rating on the follow up manipulation check question, were excluded from subsequent statistical analyses. I consider some interesting differences in the proportions of participants who noticed star ratings versus Likes below.

**Differences in ability to recall stars versus Likes.** In focus group studies, individuals have explicitly stated that they rely on online ratings when making credibility evaluations (e.g., Metzger et al., 2010). This suggests that individuals pay a lot of attention to ratings when deciding whether to trust a message. Study 1 enables me to use quantitative data to examine if this is indeed true. If most participants pay attention to online ratings, then an overwhelming majority should report that they notice the online rating associated with the fact checking message.
As noted above, research has also shown that the type of visual stimuli influences message response (Fu, 2012). This suggests that participants might differ in their ability to recall star ratings versus Likes. If participants pay equal attention to star ratings and Likes, equal percentage proportions of participants should be able to accurately recall the star ratings and Likes associated with the fact checking messages.

I ran frequency analyses to compare how participants respond to the two types of ratings in this study. Participants who correctly identify the rating condition they are assigned to were coded as ‘1 = Identified the star rating (Likes) correctly’ and all other participants were coded as ‘0 = Did not manage to identify the star rating (Likes)’. Results show that participants respond very differently to star ratings versus Likes. Although roughly equal proportions of participants notice the star ratings (83%) and Likes (81%), participants are able to recall the star rating far more accurately (64%) than the number of Likes (39%).

From the above, participants appear to exhibit differences in the way they process star ratings which are presented in the form of percentage proportions and Likes which are presented as numbers. I will return to the significance of this pattern in the discussion.

Next, I examine if there is systematic variation between participants who notice online ratings and those who do not on demographic traits. I also gauge whether participants who correctly identified the online rating associated with the fact checking messages differed from those who failed to do so accurately. As these are exploratory analyses, not theorized a priori, I examine if participants who notice
online ratings differ from those who do not on a battery of demographic variables: age, income, education, gender, political affiliation, ideology, frequency of political social media use, attitudes toward climate change, attitudes toward Obama running the country, and attitudes toward vaccines. I also run similar analyses to see if participants who correctly identify the online rating differ from those who do not manage to do so on these demographic characteristics. I also examine whether participants who noticed the ratings and those who correctly identify the online rating are more likely to pass the other attention filters too.

Independent sample \( t \) tests show, perhaps unsurprisingly, that those who notice the star ratings are significantly heavier social media users, \( M = 16.94, SD = 6.69 \), than those who do not notice the star ratings, \( M = 14.0, SD = 6.74, t(219) = 2.41, p < .05 \). At face value, it makes sense for participants who are heavier social media users to be more attuned to star ratings because such participants are likely to be savvier social media users. Furthermore, chi-square analyses show that participants who notice the star rating are more likely to have timings falling within the 7-40 minute time window than those who do not notice the star rating, \( \chi^2 (1, N = 234) = 4.92, p < .05 \). This finding indicates that participants who notice the star rating logically are also more likely to take sensible amounts of time to complete the study than those who do not notice the star rating.

There are no statistically significant demographic differences between participants who notice the Likes and participants who do not notice the Likes. However, chi-square analyses show that participants who notice the Likes are more
likely to correctly identify the message condition to which they are assigned to than those who do not notice the Likes, $\chi^2 (1, N = 494) = 4.88, p < .05$. In other words, the participants who notice the Likes are also more likely to pay attention to the fact checking message associated with these Likes.

There are no statistically significant demographic differences between participants who correctly identify the star rating and participants who fail to identify the star rating correctly. However, an independent sample $t$ test shows that participants who correctly identify the number of Likes associated with the fact checking message are more likely to have higher monthly household incomes, $M = 3.19$, $SD = 2.12$, than those who are not able to identify the number of Likes associated with the fact checking message, $M = 2.74$, $SD = 1.75$, $t(342) = -2.46$, $p < .05$. Also, an independent sample $t$ test shows that participants who correctly identify the number of Likes associated with the fact checking message are significantly younger, $M = 46.5$, $SD = 13.8$, than those who are not able to identify the number of Likes associated with the fact checking message, $M = 50.3$, $SD = 14.2$, $t(480) = 2.85$, $p < .01$. Given the exploratory nature of these analyses, it is likely that both these patterns are the result of chance.

**Random assignment checks.** I run analyses to ensure that the random assignment works by examining if there are any discernable demographical differences among participants across the 15 message conditions that passed the four filters described above.
One-way omnibus ANOVAs indicate that participants across conditions differ significantly in terms of frequency of political social media use, $F(14, 404) = 2.12$, $MSE = 40.8$, $p < .05$, $\eta^2 = .07$. It is important to note that random assignment minimizes demographic differences between conditions but does not entirely eliminate the possibility of demographic differences existing. It is plausible for demographic differences between conditions to still exist by chance, and coincidentally, there are significant differences in terms of social media use across conditions in this study. I include demographic controls in my statistical analyses and in particular, this variable was controlled for when testing my hypotheses.

**Differences between those who passed the four filters and those who failed.**

In addition to running tests to examine if the random assignment worked among those who passed the four filters, I run exploratory independent sample $t$ tests to gauge whether there are significant demographic differences between participants who pass all four filters and those who do not.

As described above, there are many independent sample $t$ tests run, increasing the chance of false positives. As such, any significant differences found are unlikely to be meaningful or important. In all, an independent sample $t$ test shows that participants who passed the four filters are significantly more likely to believe that vaccines were alright for healthy children, $M = 4.17$, $SD = .93$, than those who filtered out, $M = 4.03$, $SD = .96$, $t(845) = -2.14$, $p < .05$. Participants who are filtered out do not differ significantly from those who pass the four filters on the other demographic characteristics. It appears that participants who are filtered out for inattention are
more knowledgeable than those who passed the four filters, as they appear more likely to hold accurate attitudes about vaccines for healthy children.

**Manipulation Checks.**

*Online rating manipulation checks.* Study 1 tests whether participants use their existing beliefs and the ratings assigned to a fact-checking message to evaluate message trust. As such, this online rating manipulation check is crucial in confirming that participants accurately perceive the rating conditions. For the manipulation check to succeed, participants in the ‘1.5 star’ rating (61 Likes) conditions have to report seeing significantly fewer stars (Likes) than participants assigned to the ‘4.5 star’ (1159 Likes) rating conditions (refer to the Measures section on online rating favorability below for exact question wordings).

For both types of online ratings, participants’ perceptions accurately reflect the assigned ratings. Independent sample *t* tests show that participants assigned to the 1.5 star condition are significantly more likely to perceive that the star rating is unfavorable, *M* = 1.38, *SD* = .74, than those assigned to the 4.5 star condition, *M* = 3.82, *SD* = .58, *t*(147.02) = -23.1, *p* < .001. Participants assigned to the ‘61 Likes’ conditions are significantly more likely to report seeing fewer Likes, *M* = 2.20, *SD* = .82, than those in the ‘1059 Likes’ conditions, *M* = 3.08, *SD* = .54, *t*(139.7) = -8.72, *p* < .001.

*Heuristic reliability manipulation checks.* I ran manipulation checks that are identical to those in the stimuli pre-testing stage above to ensure that the heuristic reliability manipulation works. Scale means, standard deviations, and reliabilities of
the heuristic reliability scales for participants exposed to star ratings and Likes respectively are reported in the measures section below.

Among those exposed to star ratings, an independent sample $t$ test shows that participants assigned to the bias inducing condition are significantly less likely to perceive online ratings to be valid indicators of message trust, $M = 8.33$, $SD = 2.78$, than participants in the de-biasing condition, $M = 9.71$, $SD = 2.04$, $t(144.84) = -3.58$, $p < .001$.

Also, among those in the Likes conditions, an independent sample $t$ test shows that participants assigned to the bias inducing condition are significantly less likely to perceive online ratings to be valid indicators of message trust, $M = 7.94$, $SD = 2.56$, than participants in the de-biasing condition, $M = 9.09$, $SD = 2.75$, $t(203) = -3.09$, $p < .01$.

**Measures.**

**Belief consistency.** Participants were asked to gauge the accuracy of one of three statements, “Obama took more vacation days than George W. Bush,” “The MMR vaccine is likely to cause autism,” or “Nuclear power plants give out a lot of carbon dioxide,” on dichotomous true/false scales.

Next, I constructed scales gauging whether participants are exposed to belief consistent or inconsistent accurate information from the dichotomous belief accuracy statements outlined above and the fact checking message condition they were exposed to. For example, participants who believed it is false that Obama took more vacation days than Bush and were exposed to the fact checking message about Obama’s
vacation days were coded as ‘1’ = Exposure to belief consistent accurate information. By contrast, participants who believe it is true that Obama took more vacation days than Bush and were exposed to the accurate information about Obama’s vacation days were coded as ‘0’ = Exposure to belief inconsistent accurate information. Participants were fairly evenly divided between seeing belief consistent and inconsistent information about Obama’s vacation days (stars: 41% consistent, Likes: 48% consistent). I used a similar process to construct measures of belief consistency for the other two issues. There was a fairly even split between the percentage of participants who read belief consistent and inconsistent information about nuclear power plant emissions (stars: 46% consistent, Likes: 51% consistent). The last case was markedly less well balanced. An overwhelming majority of participants were exposed to belief consistent information about MMR vaccines and autism (stars: 93% consistent, Likes: 87% consistent). This shows that most Americans correctly understand that the vaccine does not cause autism. As outlined above, participants generally trusted the MMR vaccine message more than the other two messages. Furthermore, with this percentage imbalance, I tested the hypotheses once with all three issues and again with just the messages on nuclear power plant emissions and Obama’s vacation days. There are no substantively important differences in findings. Finally, I created composite belief consistency measures by summing responses on the three belief consistency items together. Overall, a narrow majority of participants were exposed to belief consistent information (stars: 59% consistent, Likes: 62% consistent). Belief confidence. After giving their responses on the dichotomous issue belief scales,
participants answered follow-up questions to gauge the certainty of their beliefs, “How certain or uncertain are you in your beliefs about the MMR vaccine causing autism/ President Obama taking fewer vacation days than President Bush/ nuclear power plants producing minimal amounts of carbon dioxide?”, and were measured on five-point scales, from 1 = Very uncertain to 5 = Very certain.

For participants in both the star ratings and Likes conditions, belief confidence measures were created by summing a) the belief confidence responses on the MMR vaccine among those who were exposed to fact checking messages about the MMR vaccine, b) the belief confidence responses on Obama’s vacation days among those assigned to that message condition, or c) the belief confidence responses on nuclear power plants among those in that message condition together. As with the items gauging belief accuracy, they answered these belief confidence measures before reading the messages with ratings. I created two separate belief confidence measures because I will be testing the effects of belief confidence among participants exposed to star ratings and Likes separately. An independent sample t test shows no statistically significant difference in belief confidence between participants in the star ratings and Likes conditions (star ratings: $M = 3.38$, $SD = 1.23$; Likes: $M = 3.35$, $SD = 1.18$).

**Online rating favorability.** After participants reported on whether they noticed the star rating or the number of Likes, they answered a follow up question on online rating favorability, “What was the star rating (number of Likes) that you just saw?”
with larger numbers indicating more favorable ratings on a four point response scale (star ratings: $M = 2.59$, $SD = 1.39$; Likes: $M = 2.70$, $SD = .80$).

*Perceived similarity with raters’ beliefs.* Four questions, “The people who rated this message have different values from me (reverse coded),” “The people who rated this message have different political beliefs from me (reverse coded),” The people who rated this message have different values from me” and “The people who rated this message have different political beliefs from me” gauged perceived belief similarity with raters. In both star rating and Likes conditions, this four item scale had highly reliable alphas (star ratings: Cronbach’s alpha = 0.77; Likes Cronbach’s alpha = 0.73). These four items are then combined together to form composite scales examining the extent to which participants agree that raters’ have similar issue beliefs as themselves (star ratings: $M = 12.10$, $SD = 2.66$; Likes: $M = 12.38$, $SD = 2.18$).

*Heuristic reliability.* Four items gauged heuristic reliability in Study 1. Three of these items were identical to the ones used in the heuristic reliability pre-test. One item “Messages with negative online ratings are more likely to be trustworthy” was reverse coded. This reverse coded item caused the initial scale reliabilities for these four items to be low among those exposed to star ratings (Cronbach alpha = 0.44) as well as those exposed to Likes (Cronbach alpha = .60). As such, in both the star ratings and Likes conditions, I dropped this item, yielding three-item heuristic reliability scales with higher values indicating greater levels of agreement that online ratings are reliable indicators of message trustworthiness (star ratings: $M = 9.01$, $SD = 2.54$; Likes: $M = 8.49$, $SD = 2.71$). These three-item scales had much better
Cronbach’s alphas (star ratings: Cronbach’s alpha = 0.85; Likes: Cronbach’s alpha = 0.88).

**Trust in online ratings.** I used five items from Sundar and Nass (2001) to gauge perceived online rating trustworthiness. They gauged the extent to which participants perceive star ratings (number of Likes) seen on the InfoTools portal to be poor (reverse coded), believable, unbiased, unfair (reverse coded), and objective indicators of message trustworthiness on five point Likert scales, with 1 = Strongly Disagree to 5= Strongly Agree. This five item scale yielded reliable Cronbach’s alphas in the star and ‘Likes conditions (star ratings: Cronbach alpha = 0.87; Likes: Cronbach alpha = 0.82). The five items were then combined to form indexes with higher values indicating higher levels of agreement that online ratings are trustworthy (star ratings: $M = 16.00$, $SD = 4.41$; Likes: $M = 14.50$, $SD = 4.00$).

**Message trust.** I used five items gauge message trust. Three items were identical to those used in the fact checking message stimulus pre-test. However, two items ‘unfair’ and ‘poor’ were reverse coded. This five item scale yielded high reliabilities in both star rating and Likes conditions (star ratings: Cronbach alpha = 0.81; ‘Likes: Cronbach alpha = 0.80). The five items were then combined to form indexes gauging overall agreement that the messages were trustworthy (star ratings: $M = 17.80$, $SD = 3.54$; Likes: $M = 17.99$, $SD = 3.14$). A full list of survey questions for Study 1 can be found in Appendix D.
Study 2: Survey experiment

**Design.** Study 2 utilized a 2 (belief consistent fact checking message versus belief inconsistent fact checking message) x 2 (Favorable online rating vs. Unfavorable online rating) between subjects survey-experiment. The survey experiment typically took between three and four minutes to complete, and it was part of a larger, omnibus survey administered by GfK on behalf of the Ohio State University’s School of Communication.

**Sample.** The 500 participants in this survey experiment came from a participant pool constructed using probability-based sampling techniques such as random-digital dialing and address-based sampling. Participants received incentives from GfK for participating in this study. More details on GfK’s sample can be found in Appendix E. The average age of participants was 46.8 years, $SD = 17.5$. 48% were male and the rest were female. 65.5% were White, 11.5% were Black, and 15.2% were Hispanic. 60% had at least high school diplomas (with 14.4% having a Bachelor’s degree or higher). In terms of political affiliation, 44% were Republicans or Republican leaning, 34.2% were Democrat or Democrat leaning, and the rest were Independents.

**Procedure.** First, I asked participants questions gauging their beliefs on several issues, including the three issues in Study 1. Next, they were exposed to one of three fact checking messages that were identical to those in Study 1 (see Stimuli, below). These fact checking messages were accompanied by star ratings identical to those in Study 1. Next, participants answered questions about a) belief similarity with...
raters and b) perceptions of the fact-checking message. Upon completing the questionnaire, participants were thanked for their participation.

**Stimuli.**

*Fact checking messages.* The fact checking messages in this study were identical to those used in Study 1.

*Online rating conditions.* Participants were randomly assigned to one of two star rating conditions identical to those in Study 1.

**Manipulation checks.**

*Online rating favorability.* The online rating favorability checks mirror those conducted in Study 1. An independent sample *t* test shows that the manipulation check worked. Participants in the ‘4.5 star’ condition are significantly more likely to perceive the message as having received more stars, *M* = 3.50, *SD* = .78, than the participants in the ‘1.5 star rating’ condition, *M* = 1.78, *SD* = 1.06, *t*(472) = -20.7, *p* < .001.

**Measures.**

*Issue beliefs.* Participants were asked to gauge the accuracy of one of three statements that have identical wording to those used in Study 1. However, I used five point response scales, “Definitely True”, “Probably True”, “Probably False”, “Definitely False”, and “Not Sure” to gauge responses instead of dichotomous true/false scales.

Next, those who believed it is “Definitely false” or “Probably false” that Obama took more vacation days than Bush and were exposed to the accurate
information about Obama’s vacation days were coded as ‘1’ = Exposure to belief consistent accurate information. By contrast, participants who believed it is “Definitely true” or “Probably True” that Obama took more vacation days than Bush and were exposed to the accurate information about Obama’s vacation days were coded as ‘0’ = Exposure to belief inconsistent accurate information. Participants who indicated that they were ‘unsure’ were excluded from the analyses since it is impossible to ascertain whether such people hold belief inconsistent or consistent views. Similar to Study 1, a roughly equal proportion of Americans (46.2%) were exposed to belief consistent information about Obama’s vacation days while the rest read belief consistent information.

I used a comparable process to construct measures of exposure to belief consistent versus inconsistent information for the other two issues. As with Study 1, there was a fairly even split between Americans who read to belief consistent information (56.7%) and belief inconsistent information (43.3%) about nuclear plants’ carbon emissions. As with Study 1, an overwhelming majority of participants (93.7%) were exposed to belief consistent information about MMR vaccines and autism. As with Study 1, I tested my hypotheses with the three issues, and then with just the messages on nuclear power plant emissions and Obama’s vacation days. There were no differences in findings. Next, I created composite belief consistency measures by summing the three dichotomous belief consistency responses outlined above together. Overall, 68% of Americans were exposed to belief consistent information.
Online rating favorability. The question and response scale gauging star rating favorability was identical to Study 1’s, $M = 2.60$, $SD = 1.28$. I also verified that participants could recall a numerical value of the star rating associated with the fact checking message (refer to Study 1’s Measures section on online rating favorability above for the question’s exact wording). Participants who refused to respond to this question were excluded from analyses.

Perceived similarity with raters’ beliefs. The four questions used to gauge similarity with raters’ beliefs were identical to Study 1’s, $M = 11.74$, $SD = 2.31$, Cronbach’s alpha $= .77$. A full list of survey questions for Study 2 can be found in Appendix F.

In sum, this chapter outlines Study 1’s design, sample characteristics, stimuli descriptions, stimuli pre-test findings, manipulation check results, data cleaning procedures, and means and standard deviations of measures. This chapter also reports study design, stimuli descriptions, manipulation check results, as well as measures used in Study 2. In the next chapter, I test my hypotheses and report these findings.
Chapter 4: Results

In this Chapter, I analyze Study 1’s data to examine a) whether issue beliefs influence perceptions of raters’ beliefs, b) whether perceptions of belief similarity with raters influence rating trust, c) how rating trust influences message trust, and d) whether belief confidence and heuristic reliability can increase rating trust. I further verify whether issue beliefs affect perceived similarity with raters’ beliefs using Study 2’s data. Results of these analyses are described in further detail below.

Testing Whether Issue Beliefs Influence Perceptions of Raters’ Beliefs

Recall that online rating favorability matters. This dissertation argues that individuals infer whether raters share their issue beliefs from the scores those raters give. Specifically, I assert that individuals perceive greater belief similarity with raters who give favorable rather than unfavorable ratings to belief consistent messages. Similarly, individuals perceive greater belief similarity with raters who give unfavorable rather than favorable scores to belief inconsistent messages. I use both studies to test these assertions.

I hypothesize that perceived belief similarity will influence rating trust: Perceived similarity with raters of belief consistent messages will increase as rating scores increase (H1a). However, perceived similarity with raters of belief inconsistent messages will decrease as rating favorability increases (H1b).

Results confirm that individuals make inferences about perceived similarity with raters’ beliefs from rating favorability, but results are not as simple as predicted. Perceived similarity with raters of belief consistent fact checking messages increases
as star rating favorability increases. Contrary to H1b, rating favorability has no effect on perceived similarity when individuals encounter star ratings of belief *inconsistent* messages. Perceived similarity with raters of belief inconsistent messages does not decrease as star rating favorability increases.

To understand if the effects of rating favorability on perceived similarity with raters’ beliefs depend on belief consistency, I use an interaction term in an ordinary least squares (OLS) regression model to test whether the effect of star rating favorability on perceived similarity with raters’ beliefs depends on belief consistency. I compute another similar interaction term to test whether the effects of Likes on perceived similarity depend on belief consistency. In the paragraphs below, I first describe whether effects of star rating favorability on perceived similarity depend on issue beliefs. Next, I describe findings in the context of Likes.

Results in Table 1 show that the effects of star rating favorability on perceived similarity with raters’ beliefs are dependent on whether individuals believed the fact checking message, \( b = .94, SE = .29, t(149) = 3.29, p < .01 \). In order to understand this significant interaction term better, I first plot a figure illustrating how the relationship between star rating favorability and perceived similarity with raters’ beliefs depends on belief consistency. As illustrated in Figure 3 below, as star ratings of belief consistent fact checking messages increase, so does perceived similarity with raters’ beliefs. However, as star ratings of belief inconsistent messages increase, perceived similarity with raters’ beliefs does not decrease.
Next, probing the interaction term enables us to examine the conditional direct effects of star rating favorability on perceived similarity when individuals are exposed to belief inconsistent and belief consistent information respectively. These conditional direct effects allow one to gauge whether the effects of the independent variable on the dependent variable depend on a specific value of the moderator (Hayes, 2013). The conditional direct effects in the analyses outlined below test whether: (H1a) Perceived similarity with raters of belief consistent messages increases as rating favorability increases and (H1b) Perceived similarity with raters of belief inconsistent messages decreases as rating favorability increases.

![Figure 3](image.png)

*Figure 3. Conditional direct effect of star rating on perceived similarity with raters’ issue beliefs among participants exposed to belief consistent fact checking messages and belief inconsistent fact checking messages in Study 1.*
The visual breakdowns of this interaction effect in Figure 3 above reflect the conditional direct effects found. Perceived similarity with raters of belief consistent messages increases as star rating favorability increases. This can be seen from the conditional direct effect, \( b = 1.04, SE = .19, t(149) = 5.59, p < .001 \). The unstandardized beta of this conditional direct effect can be derived by adding the values of the unstandardized betas of the ‘online rating’ and ‘online rating x belief consistency’ variables in Table 1 below under the ‘Star Ratings’ column. However, star rating favorability does not affect perceived similarity with raters of belief inconsistent messages. This can be seen from the non-significant conditional direct effect, \( b = .10, SE = .22, t(149) = .47, p = .65 \). The unstandardized beta of this conditional direct effect can be derived from the value of the unstandardized beta of the ‘online rating’ variable in Table 1 under the ‘Star Ratings’ column.

There are, however, important unanticipated differences in the way star ratings and Likes worked. The number of Likes does not influence perceptions of similarity with raters’ beliefs, regardless of whether participants are exposed to belief inconsistent or consistent fact checking messages. Among participants exposed to Likes, results outlined in Table 1 above showed that the effects of rating favorability on perceived similarity with raters’ beliefs does not depend on belief consistency, \( b = .60, SE = .41, t(186) = 1.48, p = .14 \).
Table 1

Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Ratings on Perceived Similarity with Raters’ Beliefs Depend on Belief Consistency in Study 1

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Star Ratings</th>
<th>Likes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived similarity with raters’ beliefs</td>
<td>b</td>
<td>SE</td>
<td>b</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political social media use</td>
<td>.06</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Ideology</td>
<td>-.10</td>
<td>.06</td>
<td>-.05</td>
</tr>
<tr>
<td>Climate change attitudes</td>
<td>.22</td>
<td>.18</td>
<td>.03</td>
</tr>
<tr>
<td>Vaccine attitudes</td>
<td>-.26</td>
<td>.20</td>
<td>.28</td>
</tr>
<tr>
<td>Obama’s performance attitudes</td>
<td>-.10</td>
<td>.22</td>
<td>-.32</td>
</tr>
<tr>
<td>Education</td>
<td>.13</td>
<td>.15</td>
<td>-.03</td>
</tr>
<tr>
<td>Age</td>
<td>-.01**</td>
<td>.02</td>
<td>.03*</td>
</tr>
<tr>
<td>Gender</td>
<td>1.16</td>
<td>.40</td>
<td>.07</td>
</tr>
<tr>
<td>Democrat</td>
<td>-.11</td>
<td>.60</td>
<td>.80</td>
</tr>
<tr>
<td>Republican</td>
<td>-.07</td>
<td>.58</td>
<td>-.03</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online rating</td>
<td>.10</td>
<td>.22</td>
<td>-.15</td>
</tr>
<tr>
<td>Belief consistency</td>
<td>-3.18***</td>
<td>.83</td>
<td>-.54</td>
</tr>
<tr>
<td>Online rating x Belief consistency</td>
<td>.94**</td>
<td>.29</td>
<td>.60</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>.33***</td>
<td></td>
<td>.15**</td>
</tr>
<tr>
<td>n</td>
<td>150</td>
<td></td>
<td>187</td>
</tr>
</tbody>
</table>


In sum, individuals infer whether raters share their beliefs based on star rating favorability. And the effects of star rating favorability on perceived similarity with raters’ beliefs depend on belief consistency. Perceived similarity with raters of belief consistent fact checking messages increases as star rating favorability increases.
However, star rating favorability has no effect on perceived similarity with raters of belief inconsistent fact checking messages. Furthermore, the effects of Likes on perceived belief similarity with raters does not depend on belief consistency.

**Study 2 results.** Having demonstrated that the effect of rating favorability on perceived similarity depends on issue beliefs in an experimental setting, I also examined if this pattern of findings holds true among the general population. Using a survey experiment with a nationally representative sample helps me to see if these findings have external validity without compromising internal validity. Once again, I examine whether the effects of star rating favorability on perceived rater similarity depend on belief consistency, and test H1a and H1b (described above).

Study 2 and Study 1’s findings mirror each other. The typical American infers whether raters share their beliefs from how the raters assess belief-relevant messages. As with study 1, study 2 shows that ratings of belief consistent messages only influenced perceived similarity with raters’ beliefs. Perceived similarity with raters’ of belief consistent messages increased as rating favorability increases. No such relationship is observed among those who read the belief inconsistent messages.

I use Study 1’s statistical analysis techniques described above to probe for interaction effects in Study 2. Results in Table 2 below show that the effects of star rating favorability on perceived similarity with raters’ beliefs were dependent on whether participants believed the fact checking message, $b = .80$, $SE = .27$, $t(207) = 2.95$, $p < .01$. 

45
Table 2

Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Ratings on Perceived Similarity with Raters’ Beliefs Depend on Belief Consistency in Study 2

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Perceived similarity with raters’ beliefs</th>
<th>b</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>.01</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.00</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.24</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Online health care use</td>
<td>-.08</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Social media use</td>
<td>.01</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Political media use</td>
<td>.03</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Political participation</td>
<td>-.12</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Environmentalist</td>
<td>.55</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>Democrat</td>
<td>.53</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Republican</td>
<td>-.38</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>Ideology</td>
<td>.14</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Star rating</td>
<td>-.02</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Belief consistency</td>
<td>-1.79*</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Star rating x Belief consistency</td>
<td>.80**</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>.17**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>208</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Similar to Study 1, as star ratings of belief consistent messages increase, so does perceived similarity with raters’ beliefs (refer to Figure 4 below). However, as the star ratings of belief inconsistent messages increase, perceived similarity with raters’ beliefs does not decrease.
Figure 4. Conditional direct effect of star rating on perceived similarity with raters’ issue beliefs among participants exposed to belief consistent fact checking messages and belief inconsistent fact checking messages in Study 2.

Furthermore, the conditional direct effect analyses reflect this visual pattern of findings. Perceived similarity with raters of belief consistent messages increases as star rating favorability increases, $b = 0.77$, $SE = .15$, $t(207) = 5.17$, $p < .001$. However, star ratings of belief inconsistent messages do not affect perceptions of similarity with raters’ beliefs. This can be seen from the conditional direct effect, $b = -.02$, $SE = .22$, $t(207) = -.11$, $p = .92$.

In sum, the prediction that perceived similarity with raters of belief consistent messages increases as star rating favorability increases receives further robust support from Study 2’s nationally representative sample. The average American is likely to
perceive greater similarity with raters who give belief consistent messages favorable star ratings. However, the average American does not perceive greater belief similarity with raters who give unfavorable rather than favorable ratings to belief inconsistent messages.

**Testing Whether Perceived Similarity Influences Rating Trust**

The next part of the model concerns the relationship between perceived similarity with raters’ beliefs and trust in online ratings. I expect trust in online ratings to increase as perceived similarity with raters’ beliefs increases (H2). This prediction receives robust support among participants exposed to fact checking messages with both star ratings and Likes.

To understand if perceived similarity with raters beliefs predicts trust in online ratings, I run an OLS regression model for participants exposed to star ratings and Likes respectively. Perceived similarity with raters’ beliefs was the independent variable, and the dependent variable was trust in online ratings. According to results in Table 3 below, perceived similarity with raters’ beliefs is indeed positively associated with trust in star ratings, $b = .73, SE = .14, t(149) = 5.42, p < .001$. Also, perceived similarity with raters’ beliefs is likewise positively associated with trust in Likes, $b = .71, SE = .12, t(186) = 5.93, p < .001$. 


Table 3

Unstandardized Betas and Standard Errors For Perceived Similarity With Raters’ Beliefs As a Predictor of Online Rating Trustworthiness in Study 1

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Trust in Online Ratings</th>
<th></th>
<th>Likes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Star Ratings</td>
<td>b</td>
<td>SE</td>
<td>b</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political social media use</td>
<td>.02</td>
<td>.06</td>
<td>.09*</td>
<td>.04</td>
</tr>
<tr>
<td>Ideology</td>
<td>.18</td>
<td>.11</td>
<td>.01</td>
<td>.07</td>
</tr>
<tr>
<td>Climate change attitudes</td>
<td>-.05</td>
<td>.32</td>
<td>.24</td>
<td>.25</td>
</tr>
<tr>
<td>Vaccine attitudes</td>
<td>-.02</td>
<td>.34</td>
<td>-.92**</td>
<td>.31</td>
</tr>
<tr>
<td>Obama’s performance attitudes</td>
<td>.27</td>
<td>.37</td>
<td>-.32</td>
<td>.28</td>
</tr>
<tr>
<td>Education</td>
<td>-.11</td>
<td>.26</td>
<td>-.36</td>
<td>.21</td>
</tr>
<tr>
<td>Age</td>
<td>-.01</td>
<td>.03</td>
<td>.04</td>
<td>.02</td>
</tr>
<tr>
<td>Gender</td>
<td>.86</td>
<td>.71</td>
<td>.62</td>
<td>.55</td>
</tr>
<tr>
<td>Democrat</td>
<td>-.46</td>
<td>1.05</td>
<td>-.13</td>
<td>.81</td>
</tr>
<tr>
<td>Republican</td>
<td>-1.94</td>
<td>1.00</td>
<td>-.50</td>
<td>.78</td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived similarity between raters’ beliefs and issue beliefs</td>
<td>.73***</td>
<td>.14</td>
<td>.71***</td>
<td>.12</td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>.25***</td>
<td></td>
<td>.26***</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>150</td>
<td></td>
<td>187</td>
<td></td>
</tr>
</tbody>
</table>

Note. Democrat coded ‘yes’, Republican coded ‘yes’. *p < .05, **p < .01, ***p < .001.

In sum, perceived similarity with raters’ beliefs affects trust in online ratings. As similarity with raters’ beliefs increased, so did trust in online ratings. These effects were observed in both star rating and Likes conditions.

Testing Whether Rating Trust Influences Message Trust

Now, we focus on factors shaping trust in fact checking messages. Prior scholarship suggests that ratings help individuals to decide whether to trust messages...
(Metzger et al., 2010). I argue, however, that there is more to this story. Ratings do not have a uniform influence on perceived trust; instead, the effect of ratings on message trust depends on rating trust. Specifically, I predict that the more individuals trust online ratings, the more positive the association between online ratings and message trust (H3a). A notable consequence of this argument is that extreme distrust of online ratings will result in a boomerang effect, creating a negative association between online ratings and message trust (H3b).

Results are consistent with these expectations, showing that trust in online ratings powerfully shapes the influence of ratings on message trust. Most importantly, the ratings have effects opposite of what we would typically expect when participants have low levels of trust in ratings.

To understand if the effects of ratings on message trust depended on rating trust, I use an interaction term in an ordinary least squares (OLS) regression model for participants exposed to star ratings and Likes respectively. Results in Table 4 below show that star rating favorability does not have a main effect on message trust. Rather, the effects of star rating favorability on message trust in online ratings is contingent on trust in the star ratings, \( b = .25, SE = .04, t(149) = 5.67, p < .001 \).

In order to understand the interaction term better, I plot a figure to illustrate how the relationship between star rating favorability and message trust depended on rating trust. As illustrated in Figure 5 below, when trust in ratings is high, message trust increases as the star rating increases. When trust in ratings is average, the star
rating does not appear to influence message trust. However, when trust in star ratings is low, message trust decreases as the star rating increases.

Table 4

Hierarchical Linear Regression Analyses With Unstandardized Betas and Standard Errors Examining Whether the Effects of Online Rating Favorability on Trust in Fact Checking Messages Depend on Rating Trust in Study 1

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Perceived trust in fact checking messages</th>
<th>Likes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Star Ratings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Initial $b$</td>
<td>$SE$</td>
</tr>
<tr>
<td>Step 1 Control Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political social media use</td>
<td>-.02</td>
<td>.05</td>
</tr>
<tr>
<td>Ideology</td>
<td>.06</td>
<td>.09</td>
</tr>
<tr>
<td>Climate change attitudes</td>
<td>.31</td>
<td>.26</td>
</tr>
<tr>
<td>Vaccine attitudes</td>
<td>.53</td>
<td>.28</td>
</tr>
<tr>
<td>Obama’s performance attitudes</td>
<td>.46</td>
<td>.31</td>
</tr>
<tr>
<td>Education</td>
<td>.40</td>
<td>.21</td>
</tr>
<tr>
<td>Age</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>Gender (1 = Men)</td>
<td>-.14</td>
<td>.58</td>
</tr>
<tr>
<td>Democrat (1 = Yes)</td>
<td>.25</td>
<td>.87</td>
</tr>
<tr>
<td>Republican (1 = Yes)</td>
<td>-.42</td>
<td>.84</td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online rating</td>
<td>.20</td>
<td>.22</td>
</tr>
<tr>
<td>Trust in online ratings</td>
<td>.16*</td>
<td>.07</td>
</tr>
<tr>
<td>Step 2</td>
<td>Perceived favorability x Trust in online ratings</td>
<td></td>
</tr>
<tr>
<td>Total $R^2$</td>
<td>.19**</td>
<td></td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>.16</td>
<td>.01</td>
</tr>
<tr>
<td>$n$</td>
<td>150</td>
<td>187</td>
</tr>
</tbody>
</table>

Note. Democrat coded ‘yes’, Republican coded ‘yes’. *$p < .05$, **$p < .01$, ***$p < .001$. ***$p < .001$. 51
Next, I probe the significant interaction term to understand the conditional direct effects of star rating favorability on message trust at various levels of star rating trust. Recall that conditional direct effects gauge the effects of the independent variable on the dependent variable at specific levels of the moderator. Per convention, I examine the effects of the independent variable on the dependent variable at high values of the moderator (1 standard deviation above the mean), average, and low values of the moderator (1 standard deviation below the mean). When trust in star ratings is high, star ratings work in predictable ways. As trusted star rating favorability increases, message trust also increases, lending support to H3a. This can be seen from the conditional direct effect, $b = 1.28, SE = .28, t(149) = 4.64, p < .001$.

When individuals have average levels of trust in star ratings, star rating favorability does not affect message trust. When trust in star ratings is low, however, star ratings work in a manner opposite of what scholars would typically expect. H3b, which posits that star rating favorability has a negative effect on message trust when trust in star ratings is low, is supported. Message trust decreases as star rating favorability increases. This is evident from the conditional direct effect, $b = -.93, SE = .28, t(149) = -3.29, p < .01$. These findings are consistent with descriptions of boomerang effects. However, as there is no control condition, we cannot conclude that message trust is lower than what it would have been without star ratings.
Figure 5. Trust in star ratings as a significant overall moderator of the relationship between favorability of star ratings and trust in fact checking messages. Values for low levels of trust in star rating represent values that are one standard deviation below the mean. Values for average levels of trust in star rating represent values that are at the mean. Values for high levels of trust in star ratings indicate values that are one standard deviation above the mean.

Among participants exposed to Likes, results outlined in Table 4 above show that the number of Likes has no main effect on message trust. Also, trust in Likes does not influence the relationship between the number of Likes and message trust, $b = .11$, $SE = .08$, $t(186) = 1.34$, $p = .18$. Irrespective of an individual’s trust in Likes, they have no significant effect on message trust.

In sum, the effects of ratings on message trust depend on rating trust. However, these effects are only observed among individuals exposed to star ratings, not Likes. When individuals trust star ratings a lot, message trust understandably increases as
rating favorability increases. When individuals distrust star ratings, message trust actually decreases as rating favorability increases. By contrast, the effect of Likes on message trust does not depend on trust in Likes.

Using Belief Confidence and Heuristic Reliability to Increase Rating Trust

So far, this study has shown that individuals process online ratings in a biased manner. The more individuals perceive that raters share their beliefs, the more such ratings are trusted. Next, I turn to examining how to increase trust in online ratings of fact checking messages. I examine whether belief confidence and reliance on the issue belief-rating trust heuristic influence the effect of perceived similarity on rating trust.

In the paragraphs below, I first test whether the positive relationship between perceived similarity with the rater and online rating trust becomes stronger as belief confidence increases (H4). I then test whether the positive relationship between perceived similarity and rating trust becomes stronger as heuristic reliance increases (H5). Next, I move on to examining these hypotheses in the context of Likes.

In the context of star ratings, results only partially support these expectations. When individuals exhibit high levels of reliance on the social affirmation heuristic, trust in star ratings indeed increases as perceived similarity increases. In contrast, perceived belief similarity with raters has no effect on star rating trust when individuals exhibit low levels of reliance on the heuristic. Also, when individuals exhibit high levels of belief confidence, perceived similarity has no significant effect on trust in star ratings.
Table 5

Unstandardized Betas and Standard Errors Examining Whether the Effect of Perceived Similarity With Raters’ Beliefs on Rating Trust Depends on Belief Confidence and Heuristic Reliability in Study 1

<table>
<thead>
<tr>
<th>Control Variables</th>
<th>Star Ratings</th>
<th>Likes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>SE</td>
</tr>
<tr>
<td>Political social media use</td>
<td>-.03</td>
<td>.05</td>
<td>.00</td>
<td>.03</td>
</tr>
<tr>
<td>Ideology</td>
<td>.19</td>
<td>.10</td>
<td>-.03</td>
<td>.06</td>
</tr>
<tr>
<td>Climate change attitudes</td>
<td>-.02</td>
<td>.30</td>
<td>.08</td>
<td>.20</td>
</tr>
<tr>
<td>Vaccine attitudes</td>
<td>.03</td>
<td>.32</td>
<td>-.53*</td>
<td>.25</td>
</tr>
<tr>
<td>Obama’s performance attitudes</td>
<td>.15</td>
<td>.35</td>
<td>-.51*</td>
<td>.22</td>
</tr>
<tr>
<td>Education</td>
<td>-.08</td>
<td>.24</td>
<td>-.14</td>
<td>.16</td>
</tr>
<tr>
<td>Age</td>
<td>.00</td>
<td>.03</td>
<td>.04*</td>
<td>.02</td>
</tr>
<tr>
<td>Gender</td>
<td>.48</td>
<td>.67</td>
<td>.74</td>
<td>.42</td>
</tr>
<tr>
<td>Democrat</td>
<td>-.07</td>
<td>.99</td>
<td>.37</td>
<td>.61</td>
</tr>
<tr>
<td>Republican</td>
<td>-2.05*</td>
<td>.93</td>
<td>.09</td>
<td>.61</td>
</tr>
<tr>
<td>Independent Variables</td>
<td>-.02</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief Confidence</td>
<td>1.03</td>
<td>1.39</td>
<td>-.38</td>
<td>1.01</td>
</tr>
<tr>
<td>Heuristic reliability</td>
<td>2.28***</td>
<td>.63</td>
<td>1.03*</td>
<td>.43</td>
</tr>
<tr>
<td>Perceived similarity</td>
<td>2.31***</td>
<td>.56</td>
<td>.38</td>
<td>.39</td>
</tr>
<tr>
<td>between raters’ beliefs and issue beliefs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief confidence x perceived similarity</td>
<td>-0.09</td>
<td>.11</td>
<td>.04</td>
<td>.08</td>
</tr>
<tr>
<td>Heuristic reliability x perceived similarity</td>
<td>-.15**</td>
<td>.05</td>
<td>-.01</td>
<td>.03</td>
</tr>
<tr>
<td>Total R²</td>
<td>.37***</td>
<td>.59***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>150</td>
<td>187</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Democrat coded ‘yes’, Republican coded ‘yes’. *p < .05, **p < .01, ***p < .001.

The test takes the form of two interaction terms in an ordinary least squares (OLS) regression model predicting rating trust. These two interaction terms allow me
to test whether effects of perceived similarity on rating trust depend on belief confidence and heuristic reliance.

Results show that the effects of perceived similarity on star rating trust does not depend on belief confidence, $b = -.09, SE = .11, t(149) = -.81, p = .42$ (and see Table 5 above). Hypothesis 4 is not supported for star ratings.

Moderation by heuristic reliability, the subject of H5, is however significant. In Table 5, above, we see that the effects of perceived similarity on star rating trust depends on heuristic reliability, $b = -.15, SE = .05, t(149) = -2.96, p < .01$.

Plotting the relationship between perceived similarity and star rating trust helps to illustrate this effect, making it easier to compare individuals primed to trust the heuristic to those primed to distrust it (Figure 6). When heuristic reliability is high, star rating trust increases as perceived similarity with raters’ beliefs increases. However, when heuristic reliability is low, star rating trust does not appear to increase as perceived similarity with raters’ beliefs increases.

Furthermore, I probe the significant interaction term, examining the conditional direct effects of perceived similarity on star rating trust at high values of heuristic reliability (1 standard deviation below the average score on the heuristic reliability scale), average, and low values of heuristic reliability (1 standard deviation above the average score on the heuristic reliability scale).
Figure 6. Heuristic reliability as a significant overall moderator of the relationship between perceived similarity with raters’ beliefs and trust in star rating. Values for ‘high levels of heuristic reliability’ represent values that are one standard deviation below the mean. Values for ‘low levels of heuristic reliability’ indicate values that are one standard deviation above the mean.

Table 6, below, shows how the effects of perceived similarity on star rating trust depend on heuristic reliability. When reliance on the social affirmation heuristic is high, star rating trust increases as perceived similarity with raters’ beliefs increases. This is evident from the significant conditional direct effects of perceived similarity on star rating trust at high levels of heuristic reliability. However, when reliance on the social affirmation heuristic is low, perceived similarity does not affect star rating trust. This is evident from the non-significant conditional direct effect of perceived similarity on star rating trust at low levels of heuristic reliability. Thus, Hypothesis 5 is supported for star ratings.
Table 6

Conditional Direct Effects of Perceived Similarity With Raters’ Beliefs On Trust in Online Ratings at Various Values of Heuristic Reliability in Study 1

<table>
<thead>
<tr>
<th>Star Ratings</th>
<th>Heuristic Reliability</th>
<th>b</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1.05***</td>
<td>0.18</td>
<td>6.00</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.68***</td>
<td>0.13</td>
<td>5.36</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.31</td>
<td>0.18</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Coefficients for conditional direct effects of perceived similarity with raters’ beliefs on trust in online ratings at various values of heuristic reliability are shown above at the mean level of belief confidence. Interaction patterns are the same at values of belief confidence that are +/- one standard deviation from the mean. Age, gender, political leanings, ideological leanings, education, frequency of political social media use, attitudes toward vaccinating healthy children, attitudes about climate change, and attitudes about Obama’s performance, served as demographic controls. Values of heuristic reliability are the mean and +/- one standard deviation from mean. ***p < .001.

Moving on to Likes, Table 5 shows that the effect of perceived similarity on rating trust did not depend on belief confidence, $b = .04$, $SE = .08$, $t(186) = .54$, $p = .54$. Nor is heuristic reliability influential in the context of Likes. Table 5 shows that the effect of perceived similarity on rating trust do not depend on heuristic reliability among those exposed to fact checking messages with Likes, $b = -.01$, $SE = .03$, $t(186) = -.30$, $p = .76$. H5 is not supported for Likes.

In sum, I examine whether the effects of perceived similarity on rating trust depend on belief confidence and heuristic reliability. Effects of perceived similarity on rating trust depend on heuristic reliability when individuals are responding to star
ratings. When reliance on the social affirmation heuristic is high, the more people perceive belief similarity with raters, the more they trust the star rating. There is no such interaction in the context of Likes. Furthermore, the effects of perceived similarity on rating trust do not depend on belief confidence in both star rating and Likes conditions.

**Testing For Differences Between Stars and Likes**

One common theme that emerged in the analyses outlined above shows that star ratings and Likes appear to elicit very different effects. Notably, the findings that a) issue belief affect perceptions of rater similarity and b) the effects of rating trust on message trust are consistently documented with star ratings, but not with Likes. Although these differences are not hypothesized a priori, they are nonetheless interesting and merit further investigation. I thus test whether the effects of star ratings and Likes are statistically different from each other.

Following the approach of testing the hypothesized two-way interactions that I describe in the above paragraphs, I use PROCESS to test whether the effects of star ratings differ from those of Likes. In addition to testing the two way interactions outlined in the above paragraphs, I specify rating type as an additional moderator variable in PROCESS. This approach allows me to concretely ascertain whether the effects observed above are of the same magnitude for star ratings and Likes.

Results show that the effects described in the above paragraphs do not differ significantly between rating type. In the above paragraphs, I outline how the effects of rating favorability on perceived rater similarity depend on issue beliefs, and that the
effects of rating favorability on message trust depend on rating trust. Though these effects appear to be more pronounced among star ratings than Likes, they are not statistically different from one another. This study cannot conclude with confidence that people respond differently to star ratings and Likes.
Chapter 5: Discussion

Political misperceptions impede democracy by causing people to make unsound political judgments. Research has suggested that online platforms can help to counter such deleterious misperceptions (Garrett & Weeks, 2013). This dissertation examines whether online ratings can help make factual corrections more persuasive. Findings show that online ratings have limited effectiveness in persuading people to trust factual corrections. People evidently try to deduce whether raters’ share their beliefs about factual corrections. They will distrust ratings that disaffirm these issue beliefs, and vice versa. These findings are in line with research suggesting that people are motivated to protect their beliefs (Taber & Lodge, 2006) and will reject belief inconsistent factual corrections (e.g., Lewandowsky et al., 2012). Regardless of whether people are exposed to offline or online content, their beliefs can bias information processing of politically divisive factual corrections and their associated ratings.

Furthermore, these findings show that rating trust affects evaluations of factual corrections. Notably, when people distrust ratings, favorable star ratings cause message distrust whereas unfavorable ratings lead people to trust messages. This finding challenges the assumption that online ratings always elicit bandwagon effects. Also, this appears to be consistent with research that has shown how distrusted messages yield boomerang effects that run counter to intended effects (Byrne & Hart, 2009). Instead of causing people to reject political misperceptions, belief discrepant ratings might exacerbate these misperceptions. As explained above, however, this
study did not incorporate a control condition. Although this is not the most rigorous test of boomerang effects, findings suggest the presence of such effects. Trusted star ratings are positively associated with message trust. Conversely, distrusted star ratings are negative associated with message trust.

However, findings show that de-biasing messages can undermine these boomerang effects by increasing trust in belief discrepant star ratings. Specifically, people who are exposed to de-biasing messages rely less on the social affirmation heuristic when evaluating star rating trustworthiness. They are less likely to conclude that star ratings from raters with dissimilar beliefs are untrustworthy. This finding mirrors previous research on how de-biasing messages serve as forewarnings that help people to keep their cognitive biases in check (Schul, 1993). It is also consistent with the suggestion that de-biasing messages can make people more receptive toward belief inconsistent factual corrections (Lewandowsky et al, 2012).

**Practical Implications**

This dissertation’s findings have important practical applications for using online ratings to influence trust in fact checking messages and, more broadly speaking, belief relevant messages. First, I explain why we should omit ratings when people have strong beliefs about message content, then I offer two recommendations that center on making online ratings more effective should people choose to use them. Specifically, I describe how de-biasing messages can help to increase rating trust in spite of strong prior beliefs, and I make the case for using star ratings instead of Likes to influence message trust.
One surprising implication of this work is that we may be better off omitting online ratings when people have strong prior beliefs about message content. This flies in the face of conventional wisdom, which suggests that online ratings are a reliable way to promote popular acceptance of an idea. Perhaps surprisingly, ratings are not very helpful when people have strong prior beliefs about a message’s subject matter. Under such circumstances, people trust ratings because they affirm issue beliefs, and not because they genuinely help them to evaluate online information. This means that people’s beliefs are the most important perceptual filter when deciding whether to trust online content. These findings are consistent with research showing that people are more likely to trust belief consistent information than belief challenging information (Taber & Lodge, 2006).

There may, however, be instances in which the use of ratings is unavoidable. For example, some websites such as Amazon.com are geared primarily toward providing collective user feedback. Such websites depend heavily on user generated content such as online ratings. Other websites such as Facebook have in-built social sharing mechanisms that intentionally use ratings to help content go viral. In these cases where it is hard to eschew ratings, there are some strategies that designers can use to increase the chances that these ratings will work as they are intended. The solution is to use de-biasing messages. This research shows that de-biasing messages can help to increase rating trust, even in the face of strong beliefs. Web practitioners might, for example, design online platforms that display de-biasing messages before users proceed to view online ratings and their associated content. These de-biasing
messages can be brief, describing how research has shown that ratings often concur with experts’ consensus on issues ratings, and can thus be considered reliable indicators of message quality. For example, web platform designers can consider pinning or sticking the de-biasing message with enlarged, boldface font at the top left corner of the webpage. This unobtrusive message display strategy will ensure that people read the de-biasing message before moving on. After reading these messages, people will be less likely to think that ratings are only trustworthy if they affirm issue beliefs. In turn, such increased levels of rating trust will cause people to rely more on them when evaluating message trustworthiness.

Also, if website designers persist on using online ratings to shape message trust, these findings suggest that they have to be mindful when selecting what ratings to use. These studies provide tentative evidence that star ratings elicit observable bandwagon and boomerang effects on message trust; Likes appear not to do so. People also tend to recall star ratings more readily than Likes. Together, these findings suggest that people associate star ratings with message trustworthiness. If web platform designers are confident that their target audience’s beliefs align with message content, or that users do not have strong prior beliefs, star ratings are a powerful way to influence perceptions of message trust. In the former case, people are likely to give favorable ratings to belief consistent message content, and these trusted favorable star ratings will elicit message trust. In the latter case, people will simply take star ratings at face value and rely heavily on them when evaluating message trustworthiness.
By contrast, findings offer no evidence that the number of Likes influences message trust. Unlike star ratings, most people found it a lot harder to accurately recall the number of Likes. These findings suggest that people do not view Likes as an endorsement of message trust. It is plausible that the number of Likes may send other types of signals, such as content popularity or issue importance. If so, the implications of these findings extend to other ratings that function in ways similar to Likes. Social networking sites such as Facebook continually evolve and expand their repertoire of online ratings. Some of Facebook’s online ratings such as their newly implemented ‘sad face’ or ‘love it’ ratings might trigger similar reactions to Likes. For example, ‘love it’ ratings on Facebook can indicate content interest. Conversely, ‘sad face’ ratings can imply that the online content features bad news that people ought to pay attention to. Consequently, web platform designers might consider pairing Likes or other similar cues, instead of star ratings, when users are evaluating messages on contentious issues. These cues are less likely to cause boomerang effects when people distrust them. However, web platforms might consider using star ratings for non-controversial content. People are less likely to challenge ratings given to non-controversial message content. In turn, such trusted star ratings will have much more powerful and observable bandwagon effects on message trust than Likes.

**Theoretical Implications**

Ratings have complex effects: They can either persuade or backfire, depending on circumstances that I describe in detail below. As outlined above, this challenges the assumption that ratings elicit uniformly persuasive and powerful
bandwagon effects (e.g., Messing & Westwood, 2012; Metzger et al., 2010; Sundar, et al., 2007), and it has implications for how we talk about media effects more generally. Before describing these implications, I give a brief overview of the media effects debate.

The media effects debate has largely been framed in terms of whether media have powerful or limited effects (Neuman & Guggenheim, 2011; Scheufele & Tewksbury, 2007). In the early 1930s to 1940s, scholars believed that media messages were ‘magic bullets’ that could instantaneously transform attitudes and behaviors (Lasswell, 1930). However, voter behavior studies debunked this perspective on media effects in the late 1940s (Lazersfeld, Berelson, & Gaudet, 1948). These studies showed that campaign messages had limited effects on belief change. Rather, people tended to reinforce their political beliefs by attending to belief consistent campaign messages (Lazarsfeld et al., 1948). In the 1970s, the ‘powerful media effects’ paradigm gained popularity once again. Scholars argued that media messages were extremely persuasive at shaping worldviews (Gerbner, 1998) perceptions of the opinion climate (Noelle-Neumann, 1974), and issue salience (McCombs & Shaw, 1972). In recent years, the idea of limited media effects has resurfaced. With a plethora of media channels available, coupled with political polarization of the electorate, people are even more likely to reinforce political beliefs by seeking out belief consistent content (Bennett & Iyengar, 2008).

As described above, the findings reported in this dissertation on the persuasive effects of online ratings do not fit cleanly into the ‘powerful or limited effects’
dichotomy. Rather, they show that ratings can have both powerful and limited effects at the same time. This runs counter to the idea that limited effects morph into powerful effects over time (Neuman & Guggenheim, 2011). Two factors, namely, a) audiences’ existing beliefs, and b) rating type (e.g., star ratings or Facebook Likes) might determine when ratings have powerful or limited effects on message trust.

Ratings elicit powerful bandwagon effects when they affirm beliefs. People trust ratings that affirm beliefs. A trusted favorable rating leads people to trust a message, and vice versa. A distrusted rating, however, is likely to backfire such that positive ratings actually cause message distrust, and negative ratings cause message trust.

Also, the type of rating may determine the nature of effects. Powerful bandwagon effects described above appear more pronounced among trusted star ratings than Likes. At the same time, boomerang effects described above appear to be much more noticeable among distrusted star ratings than Likes. As explained above, however, these findings must be interpreted with caution. More robust statistical analyses show that people do not differ significantly in their responses toward star ratings and Likes. Rather, this study provides tentative evidence that these differences might exist.

In sum, this dissertation’s findings show that we need a more nuanced approach in understanding media effects. Instead of asking whether media have effects, we ought to ask when media cues such as online ratings have powerful versus limited effects. Both powerful and limited media effects can occur concurrently,
depending on certain boundaries that I describe in detail below. This perspective is in line with scholars who have stressed the importance of identifying boundary conditions for media effects (e.g., Neuman & Guggenheim, 2011). Broadly speaking, my findings suggest that two boundary factors, namely, a) audience predispositions such as beliefs, attitudes or values, and b) social environmental cues such as rating type, polls, or comments, can determine when media have powerful or limited effects. Media messages that agree with audience predispositions are likely to be powerfully persuasive, and vice versa. Easily recallable social environmental cues (e.g., star ratings, eye catching visuals) are more likely to influence message persuasiveness than unrecallable cues. This is further consistent with previous works that have specifically cited these aforementioned boundary factors as moderating variables that influence media effects’ sizes (Klapper, 1960). Lastly, these boundary conditions have implications for the kinds of conclusions we can draw during theory testing. We should not hastily conclude that media have limited effects when studies yield non-significant main effects. Rather, we should test whether powerful and limited effects vary by these moderating variables described above before jumping to conclusions.

**Directions for Future Research**

This dissertation’s findings raise a few directions for future research, which I describe below. In this dissertation, I focused on examining when ratings influence trust in fact checking messages. However, message trust does not necessarily result in belief, attitudinal, or behavioral change. For example, online ratings might lead to perceptions of message credibility, but message credibility perceptions do not
necessarily predict belief accuracy. As described in further detail below, having accurate issue beliefs and performing pro-social behaviors are prized as tangible and desirable penultimate outcomes across contexts. Consequently, scholars have devoted considerable effort to understand when media messages will elicit such positive outcomes. It is thus important to move beyond message trust and examine when online ratings, and more broadly speaking social media cues, are powerful at influencing other types of outcome variables such as beliefs, attitudes, or behaviors.

The first future research angle focuses on using social media cues to counter issue misperceptions. Political misperceptions persist because people have intense negative emotions toward belief threatening political facts. These emotions are likely to cause people to dismiss such facts (Weeks, 2015). As mentioned above, though social media platforms can disseminate misperceptions, they also have the potential to counter such misperceptions. Future studies can examine how fact checking messages and social media cues work together to reduce negative emotions toward belief threatening political facts and improve political belief accuracy.

Next, future studies can examine the role social media cues play in influencing voting behavior. Voting behavior is viewed as a crucial election campaign outcome with direct implications for policy making and good governance (Bartels, 2000). Much research has sought to identify the factors influencing vote choice (e.g., Bartels, 2000; Huckfeldt & Sprague, 1995; Weeks & Garrett, 2014). In this digital era, social media cues such as ratings or comments often accompany online campaign messages (Girish & Williams, 2010). These cues can serve as word of mouth indicators about
candidate popularity, and either undermine or enhance campaign success by affecting voting behaviors. Future studies can examine when social media cues and political campaign messages successfully work in tandem to reinforce candidate attitudes and shape voting behaviors.

Future research can also examine factors influencing the persuasiveness of social media cues in non-political contexts such as health communication. In health contexts, scholars often try to understand what motivates people to adopt good health behaviors which have long-term benefits for physical well-being (Dutta-Bergman, 2005; Rice & Atkin, 2001). For example, many studies have examined how to deter adolescents and young adults from engaging in risky health practices such as unprotected sexual relations or binge drinking (e.g., Chia, 2006; Rimal & Real, 2005). Research has shown that perceptions of peer norms spur adolescents to engage in such risky health behaviors (Rimal & Real, 2003, 2005). Social media platforms are extremely popular among adolescents and young adults (Valkenburg, Peter, & Schouten, 2006). And cues on such platforms might serve as indicators of peer norms. As such, we can examine when social media cues and health campaign messages influence norm perceptions and intentions to adopt good health practices among adolescents and young adults.

Lastly, social media platforms can help to generate free publicity or awareness about future product releases (Colliander & Dahlen, 2011). However, such publicity can either backfire or succeed in the face of existing brand attitudes. For example, some movies within a popular franchise have performed very well at the box office
despite receiving unfavorable user reviews. This suggests that existing brand goodwill can affect the persuasiveness of social media publicity. Future research can examine and identify the conditions under which brand attitudes impede the effects of social media cues on product purchase intentions.

**Conclusion**

This dissertation examines whether online ratings can increase the persuasiveness of factual corrections. Specifically, I use two studies to gauge whether a) people use their beliefs to decide whether to trust ratings of factual corrections, b) rating trust influences message trust, and c) de-biasing messages and belief confidence increase rating trust.

Contrary to the notion that online ratings are uniformly persuasive, this dissertation shows that cognitive biases can undermine these ratings’ bandwagon effects. Instead of passively relying on online ratings, people will trust online ratings that affirm their political beliefs and distrust belief discrepant ratings. Belief discrepant ratings will even trigger boomerang effects when people evaluate factual correction trustworthiness.

The more things change, the more things stay the same. No matter how online platform features evolve, these features do not necessarily trump cognitive biases. Rather, these cognitive biases affect the way people evaluate such features. Furthermore, such features might even backfire when people distrust them. Moving forward, it is important to identify conditions under which online platform features reinforce cognitive biases and when they overcome such biases.
References


http://dx.doi.org/10.1037/h0046049


DOI: 10.1111/j.1540-5907.2006.00214.x


Appendix A: Study 1’s De-biasing & Bias Inducing Messages
Instructions

Read the following statement carefully. You can take as much time as you want.

You are about to view an image of an online information portal containing information on an important issue. The information will also be accompanied by star ratings, for example ★★★★★. These online ratings represent the collective views of other people and are found on many types of websites.

Research has confirmed what many people already intuitively know: online ratings from other users often do not match experts’ views on the issue. In other words, online ratings are not good indicators of trustworthiness.

After reading through this statement, please click the ‘>>’ button to see the information portal and answer some questions about it.

**For stimulus pre-test participants, instead of being instructed to click on the “Next” button to proceed with the study, the last sentence will read as “After reading through this statement, please click the ‘Next’ button to answer some questions”.

Bias inducing condition (star ratings)
De-biasing condition (star ratings)

**Instructions**

Read the following statement **carefully**. You can take as much time as you want.

You are about to view an image of an online information portal containing information on an important issue. The information will also be accompanied by star ratings, for example ⭐⭐⭐⭐⭐. These online ratings represent the collective views of other people and are found on many types of websites.

**Research has confirmed what many people already intuitively know: online ratings from other users often match experts’ views on the issue. In other words, online ratings are good indicators of trustworthiness.**

After reading through this statement, please click the ‘>>’ button to see the information portal and answer some questions about it.
Instructions

Read the following statement carefully. You can take as much time as you want.

You are about to view an image of an online information portal containing information on an important issue. The information will also be accompanied by Likes, for example 🎉. These online ratings represent the collective views of other people and are found on many types of websites.

Research has confirmed what many people already intuitively know: online ratings from other users often do not match experts’ views on the issue. In other words, online ratings are not good indicators of trustworthiness.

After reading through this statement, please click the ‘>>’ button to see the information portal and answer some questions about it.
De-biasing condition (Likes)

**Instructions**

Read the following statement *carefully*. You can take as much time as you want.

You are about to view an image of an online information portal containing information on an important issue. The information will also be accompanied by Likes, for example 👍. These online ratings represent the collective views of other people and are found on many types of websites.

**Research has confirmed what many people already intuitively know: online ratings from other users often match experts’ views on the issue. In other words, online ratings are good indicators of trustworthiness.**

After reading through this statement, please click the ‘>>’ button to see the information portal and answer some questions about it.
Appendix B: Study 1 & Study 2’s Fact Checking Messages and Online Ratings
The MMR Vaccine and Autism

10,123 people viewed the information below.

**Question:** Is the MMR vaccine likely to cause autism?

**Answer:** There is no scientifically proven link between the MMR vaccine and autism. Reports from both the American Academy of Pediatrics and the Centers for Disease Control and Prevention conclude that there is no proven association between the MMR vaccine and autism.

Autism is a chronic developmental disorder, often first identified in toddlers from age 18 months to 30 months. The MMR vaccine is administered just before the peak age of onset of autism symptoms. This timing leads some parents to mistakenly assume a causal relationship. There is no evidence that the MMR vaccine causes autism.
The MMR Vaccine and Autism

1059 people like this

10,123 people viewed the information below.

**Question:** Is the MMR vaccine likely to cause autism?

**Answer:** There is no scientifically proven link between the MMR vaccine and autism. Reports from both the American Academy of Pediatrics and the Centers for Disease Control and Prevention conclude that there is no proven association between the MMR vaccine and autism.

Autism is a chronic developmental disorder, often first identified in toddlers from age 18 months to 30 months. The MMR vaccine is administered just before the peak age of onset of autism symptoms. This timing leads some parents to mistakenly assume a causal relationship. There is no evidence that the MMR vaccine causes autism.
The MMR Vaccine and Autism

Question: Is the MMR vaccine likely to cause autism?

Answer: There is no scientifically proven link between the MMR vaccine and autism. Reports from both the American Academy of Pediatrics and the Centers for Disease Control and Prevention conclude that there is no proven association between the MMR vaccine and autism.

Autism is a chronic developmental disorder, often first identified in toddlers from age 18 months to 30 months. The MMR vaccine is administered just before the peak age of onset of autism symptoms. This timing leads some parents to mistakenly assume a causal relationship. There is no evidence that the MMR vaccine causes autism.
Q: Who took more vacation days, President Obama or President Bush?

A: According to CBS News White House Correspondent Mark Knoller, who has covered every president since Gerald Ford and tracks the commander in chief’s travel, President Obama took fewer vacation days than President Bush. On August 8, the day before President Obama left for Martha’s Vineyard, Knoller tweeted that President Obama had spent 125 full or partial days on vacation. At the same point in President Bush’s presidency, he had spent 381 days at his Texas ranch plus 26 days at his parents’ home in Kennebunkport, Maine, for a total of 407.
Text of the Obama’s Vacation Days Fact Checking Message

Obama’s Vacation Days

Question: Who took more vacation days, President Obama or President Bush?

Answer: Considerable evidence indicates that President Obama took fewer vacation days than President Bush. According to CBS News White House Correspondent Mark Knoller, who has covered every president since Gerald Ford and tracks the commander in chief’s travel, it is false that President Obama took more vacation days than President Bush. This fact has also been confirmed by several non-partisan news sources, such as USA TODAY and PBS.

On August 8th, the day before President Obama left for Martha’s Vineyard, President Obama had spent 125 full or partial days on vacation. At the same point in President Bush’s presidency, he had spent 381 days at his Texas ranch plus 26 days at his parents’ home in Kennebunkport, Maine, for a total of 407.
Nuclear Plant Emissions and Carbon Dioxide

**Question:** Do nuclear power plants emit a lot of carbon dioxide?

**Answer:** Scientific evidence indicates that nuclear power plants emit minimal levels of carbon dioxide. All nuclear facilities are carefully monitored by the Nuclear Regulatory Commission and state regulators. According to the Nuclear Energy Institute (NEI) and research done at the National Renewable Energy Laboratory, carbon dioxide emissions from nuclear power plants are among the lowest of any electricity generation source.

The NEI reports that nuclear power plants emit 17 tons of carbon dioxide equivalent per gigawatt-hour. This level of carbon dioxide emissions is comparable to clean energy sources such as geothermal energy (15 tons per gigawatt hour) and wind energy (14 tons per gigawatt hour). In 2013, nuclear energy facilities prevented 588.5 million metric tons of carbon dioxide emissions, equal to the amount of carbon dioxide emissions from 110 million cars.
Appendix C: Stimulus Pre-Test Questions
Stimulus Pre-Test of Experimental Stimuli with Fact Checking Messages
Participants who are assigned to view either one of three issues with accurate information during the stimulus pre-test will be asked the following questions pertaining to message trust after reading the information.

1. “The message that you just read was accurate.”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)

2. “The message that you just read was believable.”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)

3. “The message that you just read was fair.”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)

4. “The message that you just read was unbiased.”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)

5. “The message that you just read was objective.”
Also, participants who are evaluating the persuasiveness of the factual content will be asked the following questions, depending on the type of issue they read:

“The information that you just read said that Obama had _______ vacation days than Bush.”

2 = More
1 = Fewer
0 = Don’t know/refused (not shown)

“The information that you just read said that the MMR vaccine is ______ to cause autism”

2= Likely
1= Unlikely
0 = Don’t know/refused (not shown)

“The information that you just read said nuclear power plants emit _______ carbon dioxide”

2= A lot of
1= Minimal amounts of
0 = Don’t know/refused (not shown)

Stimulus Pre-test to Examine Whether the Heuristic Reliability Induction Works
Participants who are tasked with evaluating whether the heuristic reliability induction works will be asked to give their responses to the statements below after viewing either the de-biasing message or the bias inducing message:

1. “Online ratings by individuals match the collective views of experts.”
   5 = Strongly Agree
4 = Agree
3 = Neither agree nor disagree
2 = Disagree
1 = Strongly Disagree
0 = Don’t know/refused (not shown)

2. “Online ratings are valid indicators of message trustworthiness”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)

3. “Online ratings are often reliable indicators of message trustworthiness”
   5 = Strongly Agree
   4 = Agree
   3 = Neither agree nor disagree
   2 = Disagree
   1 = Strongly Disagree
   0 = Don’t know/refused (not shown)
Appendix D: Study 1’s Survey Questions
Beliefs

Please indicate whether you think this statement is true or false: “The MMR vaccine is likely to cause autism.”
- True (1)
- False (2)

Please indicate whether you think this statement is true or false: “President Obama took fewer vacation days than President Bush.”
- True (1)
- False (2)

Please indicate whether you think this statement is true or false: “Nuclear power plants produce minimal amounts of carbon dioxide.”
- True (1)
- False (2)

[Randomized Question Order]

Belief Confidence

How certain or uncertain are you in your beliefs about the MMR vaccine causing autism?
- 1 Very Uncertain (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 Very Certain (5)
How certain or uncertain are you in your beliefs about President Obama taking fewer vacation days than President Bush?
- 1 Very Uncertain (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 Very Certain (5)

How certain or uncertain are you in your beliefs about nuclear power plants producing minimal amounts of carbon dioxide?
- 1 Very Uncertain (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 Very Certain (5)

**Likes Manipulation Check Questions**

Did you notice the Likes attached to the message on the InfoTools portal that you just read?
- Yes (1)
- No (2)

If you answered 'Yes' to the question above, about how many Likes did the message receive?
- 10 (1)
- 100 (2)
- 1,000 (3)
- 10,000 (4)
- There was a rating but I don't know what it said (5)

**Star Rating Manipulation Check Questions**
Did you notice the star rating attached to the message on the InfoTools portal that you just read?
☐ Yes (1)
☐ No (2)

If you answered 'Yes' to the question above, what was the star rating that you just saw?
☐ 1.5 star rating (1)
☐ 2.5 star rating (2)
☐ 3.5 star rating (3)
☐ 4.5 star rating (4)
☐ There was a rating but I don't know what it said (5)

Data Cleaning Measure on Message Identification

We want to know what you remember about the message posted on the InfoTools portal. Was this message about:
☐ Obama's Vacation Days (1)
☐ The MMR Vaccine (2)
☐ Nuclear Power Plants (3)

[Randomized Question Order]

Perceived Belief Similarity with Raters

We’d also like to know what you think about the people who submitted the rating shown on the InfoTools portal. Please indicate how strongly you agree or disagree with the following statements. The people who rated the message on the InfoTools portal:
<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither Agree nor Disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have different values from me. (1)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Have different political beliefs from me. (2)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Are similar to me. (3)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Have similar political leanings as me. (4)</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>
**Rating Trustworthiness**

We’d like to know what you think about the star rating (the number of Likes) that you just saw on the InfoTools portal. Please indicate how strongly you agree or disagree with the following statements.  

<table>
<thead>
<tr>
<th>Description</th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither Agree nor Disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A poor indicator of message trustworthiness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
</tr>
<tr>
<td>A believable indicator of message trustworthiness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
</tr>
<tr>
<td>A unfair indicator of message trustworthiness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
</tr>
<tr>
<td>An unbiased indicator of message trustworthiness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
</tr>
<tr>
<td>An objective indicator of message trustworthiness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
<td>○</td>
</tr>
</tbody>
</table>
[Randomized Question Order]
Heuristic Reliability
Lots of messages receive online ratings from other users, such as Facebook posts, news stories, or even book reviews. Now, please think about the trustworthiness of online ratings more generally. Please consider star ratings, Likes, and any other online rating that you have seen on the web when answering these questions:

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither Agree nor Disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online ratings are valid indicators of message trustworthiness. (1)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Online ratings are reliable indicators of message trustworthiness. (2)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Messages with positive online ratings are more likely to be trustworthy. (3)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Messages with negative online ratings are more likely to be trustworthy. (4)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
[Randomized Question Order]

Trust in Fact Checking Messages

We also want to know what you think about the message on the InfoTools portal. Please indicate how strongly you agree or disagree with the following statements. The message that you just read on the InfoTools portal was:

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neither Agree nor Disagree (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inaccurate (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believable (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfair (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbiased (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Demographics

What is your gender?
- Male (1)
- Female (2)

To begin, we have a few questions about how you use the Internet. In the past one month, how often did you use an online search engine to help you find information on the Internet?
- Never (1)
- Rarely (2)
- Several times a month (3)
- Several times a week (4)
- Every day or almost every day (5)
In the past month, which social networking sites did you use? Please check all that apply.

- Facebook (1)
- Google Plus (2)
- Instagram (3)
- LinkedIn (4)
- Myspace (5)
- Twitter (6)
- Vine (7)
- Others please specify (8) ____________________
- I did not use social media in the past month. (9)
PolSNS How often did you use social networking sites to do the following in the past month?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Never (1)</th>
<th>Rarely (2)</th>
<th>Several times a month (3)</th>
<th>Several times a week (4)</th>
<th>Every day or almost every day (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read news headlines or short news summaries.</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>View an online news video.</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Read political opinions.</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Read messages from politicians.</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Post, forward, or comment on anything related to politics, including</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>news stories, opinions, images, or videos.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share news you viewed automatically, without any extra clicks, using a</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>sharing service such as the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Issue Attitudes
Humans are the primary cause of climate change.
- Strongly Disagree (1)
- Disagree (2)
- Neither Agree nor Disagree (3)
- Agree (4)
- Strongly Agree (5)

Generally speaking, vaccines are safe for healthy children.
- Strongly Disagree (1)
- Disagree (2)
- Neither Agree nor Disagree (3)
- Agree (4)
- Strongly Agree (5)

President Obama is doing a good job of running the country.
- Strongly Disagree (1)
- Disagree (2)
- Neither Agree nor Disagree (3)
- Agree (4)
- Strongly Agree (5)

Political Identification
Generally speaking, when it comes to political parties in the United States, how would you best describe yourself?
When thinking about politics in general, how would you describe your views?
- Very Liberal (1)
- Liberal (2)
- Somewhat Liberal (3)
- Moderate (4)
- Somewhat Conservative (5)
- Conservative (6)
- Very Conservative (7)

Please take some time to answer the following questions.

When thinking about economic issues in general, how would you describe your views?
- Very Liberal (1)
- Liberal (2)
- Somewhat Liberal (3)
- Moderate (4)
- Somewhat Conservative (5)
- Conservative (6)
- Very Conservative (7)
When thinking about social issues in general, how would you describe your views?
- Very Liberal (1)
- Liberal (2)
- Somewhat Liberal (3)
- Moderate (4)
- Somewhat Conservative (5)
- Conservative (6)
- Very Conservative (7)

What is your current age? Please use whole numbers, e.g., 35.

What is the highest grade in school, or level of education that you’ve completed and received credit for?
- Eighth grade or less (1)
- Some high school (2)
- High school graduate or G.E.D. (3)
- Technical, trade, or vocational school after high school (4)
- Some college, no 4-year degree (including 2 year associate degree) (5)
- College graduate (B.S., B.A., or other 4-year degree) (6)
- Post-graduate training or professional schooling after college (for example, toward a Master’s Degree or Ph.D.; law or medical school) (7)

When describing your race, would you best describe yourself as:
- Asian/Asian American or Pacific Islander (1)
- Black or African-American (2)
- Native American (3)
- White/European American (4)
- Something else, please specify (5) ____________________

Are you Hispanic or Latino?
- Yes, I am Hispanic or Latino (1)
- No, not Hispanic or Latino (2)

Lastly, what is your monthly household income?
- Less than $20,000 (1)
- $20,000 to $39,999 (2)
- $40,000 to $59,999 (3)
- $60,000 to $79,999 (4)
- $80,000 to $99,999 (5)
- $100,000 to $119,999 (6)
- $120,000 to $139,999 (7)
- $140,000 to $159,999 (8)
- $160,000 to $179,999 (9)
- $180,000 or more (10)
Sample Overview

KnowledgePanel®, created by GfK, is a “probability-based online non-volunteer access panel. Panel members are recruited using a statistically valid sampling method. KnowledgePanel’s published sample frame of residential addresses covers approximately 97% of US households. When recruited, sampled noninternet households receive a netbook computer and free internet service, so that they may also participate as online panel members. KnowledgePanel consists of about 55,000 adult members (ages 18 and older) and includes persons living in cell phone-only households. The multi-dimensional Hispanic population is also represented in KnowledgePanel with members recruited in both English and Spanish, thereby representing different levels of language proficiency and acculturation levels (GFK, 2012, p. 1).”

Address Based Sampling Method

GFK participants are recruited using addressed based sampling methods (ABS). ABS “involves probability-based sampling of addresses from the US Postal Service’s Delivery Sequence File. Randomly sampled addresses are invited to join KnowledgePanel through a series of mailings (English and Spanish materials) and by telephone follow-up to non-responders when a telephone number can be matched to the sampled address. Invited households can join the panel by one of several means: completing and mailing back an acceptance form in a postage-paid envelope; calling a toll-free hotline staffed by bilingual recruitment agents; or going to a dedicated recruitment website and completing the recruitment information online. The address sampling, conducted throughout the year, is done without replacement. Addresses with matched telephone numbers from the former RDD recruitment samples (for the last five years of calling) are also removed to eliminate duplication (GFK, 2012, p. 2).”

Panel Survey Sampling

“Once panel members are profiled, they become “active” for selection for specific surveys. Samples are drawn from among active members using a probability proportional to size (PPS) weighted sampling approach. Customized stratified random sampling based on profile data is also conducted, as required by specific studies […]The selection methodology, which KnowledgePanel has used since 2000, assures
that multiple sequential samples from a finite panel membership will each reliably represent the US population (GFK, 2012, p. 3).”

Response Rates

“The within survey response rate for KnowledgePanel is 65% with some minor variation, depending on survey length and topic. In contrast, non-probability online panels typically achieve a survey response rate in the 2% to 16% range. As a result it is important to note that gross panel size is an incomplete measure of panel “scalability.” The effective panel size also depends on this within survey response rate. For example, in order to achieve 1,200 completed surveys, GfK requires only about 1,850 panelist invitations, compared to a starting sample of 50,000+ invitations for many non-probability panels. KnowledgePanel’s size of approximately 50,000 is comparable to an effective opt-in panel size of a 1.5 million (assuming a 2% within survey response rate) (GFK, 2012, p. 3-4).”

Statistical Weighting

“KnowledgePanel® sample begins as an equal probability sample that is self-weighting with several enhancements incorporated to improve efficiency. Since any alteration in the selection process is a deviation from a pure equal probability sample design, statistical weighting adjustments are made to the data to offset known selection deviations. These adjustments are incorporated in the sample’s base weight. There are also several sources of survey error that are an inherent part of any survey process, such as non-coverage and non-response due to panel recruitment methods and to inevitable panel attrition. These sources of sampling and non-sampling error are addressed using a panel demographic poststratification weight as an additional adjustment (GFK, 2012, p. 3-4).”
Appendix F: Study 2’s Survey Questions
Next we’d like to know what you believe about a few different topics.

[RANDOMIZE ORDER OF Q1, 2 & 3.]

1. Please indicate whether you think this statement is true or false:
   “President Obama took more vacation days than President Bush.”
   1. Definitely True (4)
   2. Probably True (3)
   3. Probably False (2)
   4. Definitely False (1)
   5. Not Sure (0)

2. Please indicate whether you think this statement is true or false:
   “The MMR vaccine is likely to cause autism.”
   1. Definitely True (4)
   2. Probably True (3)
   3. Probably False (2)
   4. Definitely False (1)
   5. Not Sure (0)

3. Please indicate whether you think this statement is true or false:
   “Nuclear power plants give out a lot of carbon dioxide.”
   1. Definitely True (4)
   2. Probably True (3)
   3. Probably False (2)
   4. Definitely False (1)
   5. Not Sure (0)

Now please take a moment to answer a few questions about the message that you just viewed.

4. The star rating that you saw just now was:
1. ★★★★★ (1.5 star rating) 1
2. ★★★★☆ (2.5 star rating) 2
3. ★★★☆☆ (3.5 star rating) 3
4. ★★★☆☆ (4.5 star rating) 4

**[DISPLAY ON THE SAME PAGE]**

Now please tell us how much you agree or disagree with each of these statements about the people who rated the message on the InfoTools website.

<table>
<thead>
<tr>
<th>[RANDOMIZE ORDER]</th>
<th>Strongly Disagree 5</th>
<th>Disagree 4</th>
<th>Neutral 3</th>
<th>Agree 2</th>
<th>Strongly Agree 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. The people who rated this message have different values from me.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>6. The people who rated this message have different political beliefs from me.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>7. The people who rated this message are similar to me.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>8. The people who rated this message have similar political leanings as me.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>