

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Abeywickrama, Ravini Savindya

Title:

Pro-attitudinal advocacy, depolarization, and communicative intentions: The role of self-persuasion

Date:

2020

Persistent Link:

<https://hdl.handle.net/11343/274840>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

**Pro-Attitudinal Advocacy, Depolarization, and Communicative Intentions: The Role of
Self-Persuasion**

RAVINI SAVINDYA ABEYWICKRAMA

ORCID ID: 0000-0002-8588-8402

Submitted in Total Fulfilment of the Doctor of Philosophy

November 2020

Melbourne School of Psychological Sciences

Faculty of Medicine, Dentistry and Health Sciences

Supervised by:

Dr. Simon M. Laham

Melbourne School of Psychological Sciences

Prof. Yoshi Kashima

Melbourne School of Psychological Sciences

Table of Contents

LIST OF FIGURES	VII
LIST OF TABLES	VII
DECLARATION	VIII
PREFACE	IX
<i>Contributions By Others</i>	<i>ix</i>
PUBLICATIONS INCLUDED IN THESIS	X
OTHER PUBLICATIONS DURING CANDIDATURE	XI
ACKNOWLEDGEMENTS	XII
ABSTRACT	XIV
CHAPTER 1 INTRODUCTION.....	1
A REVIEW OF “STANDARD” PERSUASION.....	3
A REVIEW OF SELF-PERSUASION.....	5
<i>Group Polarization</i>	6
<i>Mere Thought</i>	7
<i>Saying-is-Believing</i>	7
<i>Self-Reflection</i>	8
<i>Counter-Attitudinal Advocacy</i>	8
PRO-ATTITUDINAL ADVOCACY AND SELF-PERSUASION	9
<i>Meta-Cognitive Confidence, Elaboration, and Self-Persuasion.</i>	11
<i>Manipulations of PAA</i>	12
Moral vs Practical.....	12
Attack vs Defend.	14
THE CURRENT RESEARCH	15
CHAPTER 2 HISTORY OF ATTITUDES & ATTITUDE CHANGE.....	18
<i>Traditional Theories of Persuasion</i>	20
Attitude change processes.	23
Dual process models of persuasion.	24
SUMMARY	29
CHAPTER 3 EMERGING THEORIES OF PERSUASION.....	32
INTRODUCTION TO CHAPTER 3	32

<i>A Review of Self-Persuasion</i>	33
Group polarization.....	35
Mere thought.....	37
Saying-is-believing.....	39
Self-reflection.....	42
Counter-attitudinal advocacy.....	43
Pro-attitudinal advocacy.....	46
THE CURRENT RESEARCH	49
CONCLUDING REMARKS	53
CHAPTER 4 META-COGNITIVE CONFIDENCE, NEED-FOR-COGNITION, AND SELF-PERSUASION	54
PREFACE TO CHAPTER 4	54
INTRODUCTION.....	57
<i>Meta-Cognition and Self-Persuasion</i>	58
<i>Effects of Need-for-Cognition</i>	60
<i>The Current Research</i>	62
STUDY 1	63
<i>Method</i>	63
Participants and design.....	63
Materials and procedure	63
<i>Results and Discussion</i>	66
Time 1 attitude distribution.....	66
Attitude depolarization.....	66
<i>Concluding remarks</i>	68
STUDY 2	69
<i>Method</i>	69
Participants and design.....	69
Materials and procedure	70
<i>Results and Discussion</i>	72
Attitude depolarization.....	72
GENERAL DISCUSSION.....	77
<i>Meta-Cognitive Confidence, Depolarization and Communicative Intentions</i>	78
<i>Moderating Role of NfC</i>	80

LIMITATIONS AND FUTURE DIRECTIONS.....	82
CONCLUSIONS.....	83
SUPPLEMENTARY MATERIALS.....	84
<i>Effects of Target and Initial Stance (Study 2)</i>	87
CHAPTER 5 MORAL VS PRACTICAL: MORAL EXPRESSIVENESS AND SELF-PERSUASION	92
PREFACE TO CHAPTER 5	92
ABSTRACT	96
NON-TECHNICAL SUMMARY	97
<i>Background</i>	97
<i>Why Was This Study Done?</i>	97
<i>What Did the Researchers Do and Find?</i>	97
<i>What Do These Findings Mean?</i>	98
INTRODUCTION.....	99
THE PROBLEM WITH MORAL RHETORIC	100
PRO-ATTITUDINAL ADVOCACY AND SELF-PERSUASION.....	101
PROPOSED ALTERNATIVE: PRACTICAL REASONING	103
PROPOSED MECHANISMS.....	104
<i>Moral Foundations Theory</i>	105
<i>Deontology vs Consequentialism</i>	106
<i>Moral Emotions</i>	107
<i>Meta-Cognitive Confidence</i>	108
THE CURRENT RESEARCH	109
STUDY 1	110
<i>Method</i>	110
Participants and design.	110
Materials and procedure	110
Characterizing moral content.....	111
<i>Results and Discussion</i>	113
Differences between conditions.....	113
STUDY 2	117
<i>Method</i>	117
Participants and design.	117

Materials and procedure	117
<i>Results and Discussion</i>	119
Differences between conditions.....	119
Predictors of attitude polarization.....	121
STUDY 3	125
<i>Method</i>	125
Participants and design.	125
Materials and procedure	125
<i>Results and Discussion</i>	126
Differences between conditions.....	126
Predictors of proselytization.....	129
Mini meta-analysis	131
GENERAL DISCUSSION.....	131
<i>Moral Psychology Theory in Self-Persuasion Research</i>	133
Construct distinctiveness.	133
Importance of the Binding Moral Foundations.	134
Null Effects of Political Orientation?	135
LIMITATIONS AND FUTURE DIRECTIONS.....	137
<i>Failure to Replicate Polarization-Moral Expressiveness Link</i>	137
<i>Lack of “Control” Condition</i>	139
<i>Inability to Draw Causal Inferences</i>	139
CONCLUSIONS.....	140
FUNDING.....	142
COMPETING INTERESTS	142
DATA AVAILABILITY	142
ACKNOWLEDGEMENTS	142
SUPPLEMENTARY MATERIALS.....	143
<i>Appendix A</i>	143
Attitude issue descriptions.....	143
<i>Appendix B</i>	143
Advocacy task instructions.....	143
<i>Appendix C</i>	146
Constructing dictionaries.....	146
CHAPTER 6 ATTACK VS DEFEND: ADVOCACY STYLE AND SELF-	

PERSUASION	155
PREFACE TO CHAPTER 6	155
ABSTRACT	158
INTRODUCTION	159
SELF-PERSUASIVE EFFECTS OF PRO-ATTITUDINAL ADVOCACY	160
ATTACK VS DEFENSE ADVOCACY STYLES.....	161
ADVOCACY EFFECTS ON ATTITUDES AND COMMUNICATIVE INTENTIONS	163
HYPOTHESES	165
<i>Meta-Cognition Hypothesis</i>	165
<i>Emotion Hypothesis</i>	167
THE PRESENT RESEARCH.....	168
STUDY 1	169
<i>Method</i>	169
Participants.	169
Materials and procedure	169
<i>Results</i>	172
Data screening and cleaning.....	172
Advocacy effects on attitude polarization and meta-cognitive confidence	172
Advocacy effects on emotion.	173
Advocacy effects on communicative intentions.....	174
<i>Discussion</i>	177
STUDY 2	179
<i>Method</i>	180
Participants.	180
Materials and Procedure	180
<i>Results</i>	183
Advocacy effects on attitude polarization and meta-cognitive confidence	183
Advocacy effects on emotion.	183
Advocacy effects on accuracy and defense motivations.	183
Advocacy effects on communicative intentions.....	184
Using accuracy motivation to predict communicative intentions (motivation hypothesis).....	185
GENERAL DISCUSSION.....	188

<i>Summary</i>	188
<i>Is It Good to Attack?</i>	189
<i>Limitations and Future Directions</i>	191
Replicability & generalizability.....	191
Alternative explanations.	193
CONCLUSIONS.....	195
ACKNOWLEDGEMENTS	196
FUNDING DETAILS	196
DECLARATION OF INTEREST	196
DATA AVAILABILITY STATEMENT	196
SUPPLEMENTARY MATERIALS.....	197
<i>Study 1</i>	197
Method.....	197
Results.	200
<i>Study 2</i>	200
Method.....	200
Results	201
CHAPTER 7 GENERAL DISCUSSION	203
INTRODUCTION.....	203
THE CURRENT RESEARCH	204
<i>Theoretical Implications</i>	208
Multiple Mechanisms Underlying Self-Persuasion.....	209
Direct Impact on Communicative Intentions.	211
Re-framing Pro-attitudinal Advocacy Attempts.....	213
<i>Practical Implications</i>	215
Induce Uncertainty in People’s Arguments.....	216
Request Arguments to be Framed in Practical Terms.	217
Encourage Focus on Opponent’s Arguments.	218
<i>Limitations and Future Directions</i>	219
Replicability of Depolarization Effects.	219
Correlation vs Causation.	221
Construct Distinctiveness.	222
Generalizability to In-Person Conversations.....	222

Behaviors vs Behavioral Intentions.....	223
Using text-based analysis.	223
<i>Conclusion</i>	224
CHAPTER 8 REFERENCES.....	225

List of Figures

FIGURE 2-1	20
FIGURE 2-2	22
FIGURE 4-1	75
FIGURE 4-2	76
FIGURE 5-1	123
FIGURE 5-2	130
FIGURE 6-1	187

List of Tables

TABLE 4-1	84
TABLE 4-2	85
TABLE 4-3	86
TABLE 4-4	88
TABLE 4-5	89
TABLE 4-6	90
TABLE 4-7	91
TABLE 5-1	114
TABLE 5-2	115
TABLE 5-3	120
TABLE 5-4	121
TABLE 5-5	127
TABLE 5-6	128

Declaration

This thesis comprises my original work towards the Doctor of Philosophy (Psychology), except where indicated in the preface. Due acknowledgement has been made in the text to all other materials used. This thesis is fewer than the maximum word limit in length, exclusive of tables, figures, references and appendices, as approved by the Research Higher Degrees Committee.

Preface

This thesis is comprised of my original work, and does not contain any material previously published by another person, except where referenced in-text. Contributions of others to jointly authored works which have been included in the thesis, have been clearly stated in the Declaration for Thesis with Publication, and Co-author authorization forms, which have been submitted to the Thesis Examination System. This thesis includes my original research work undertaken since the commencement of the degree, and does not include any work required to qualify for a degree, diploma or award from any other tertiary institution, nor any work from a previous degree. Third party editorial assistance was not provided in preparation of the thesis.

Contributions By Others

Dr. Simon Laham, Dr. Damien Crone, and Mr. Joshua Rhee provided valuable input on study design, data analysis, and feedback on chapter drafts and the manuscripts included as chapters (as co-authors, where applicable). An Australian Government Research Training Program Scholarship, including stipend and fee offset was awarded for the duration of my candidature.

Publications Included in Thesis

Chapter 4:

Abeywickrama, R.S., & Laham, S. M. (2020). Meta-Cognition Predicts Attitude

Depolarization and Intentions to Engage with the Opposition Following Pro-Attitudinal

Advocacy. *Social Psychology*. <https://doi.org/https://doi.org/10.1027/1864-9335/a000424>

[Accepted for publication on 10.06.2020. Published online on 12.08.2020].

Chapter 5:

Abeywickrama, R.S., Rhee, J.J., Crone, D.L., & Laham, S. M. (2020a). Why Moral

Advocacy Leads to Polarization and Proselytization: The Role of Self-Persuasion.

Journal of Social and Political Psychology, 8(2), 473–503.

<https://doi.org/https://doi.org/10.5964/jspp.v8i2.1346>

[Accepted for publication on 28.05.2020. Published online on 02.09.2020]

Chapter 6:

Abeywickrama, R.S., & Laham, S.M. (2020b). The Effects of Attack versus Defense

Advocacy Styles on Uncongeniality Bias and Openness to Opposing Views. *Basic and*

Applied Social Psychology (in progress).

Other Publications During Candidature

Abeywickrama, R. S., Laham, S. M., & Crone, D. L. (2018). Immigration and receiving communities: The utility of threats and emotions in predicting action tendencies toward refugees, asylum-seekers and economic migrants. *Journal of Social Issues*, 74(4), 756-773. <https://doi.org/10.1111/josi.12297>

Rayner, G., Micallef, S., **Abeywickrama, R.S.,** & Wilson, S. J. (2019). Pediatric epilepsy surgery patients show normal psychosocial development at long-term follow-up despite dissatisfying family dynamics. *Epilepsy & Behavior*, 92, 245-252.
<https://www.doi.org/10.1016/j.yebeh.2019.01.004>

Le Pelley, M. E., Watson, P., Pearson, D., **Abeywickrama, R. S.,** & Most, S. B. (2019). Winners and losers: Reward and punishment produce biases in temporal selection. *Journal of experimental psychology: learning, memory, and cognition*, 45(5), 822-833. <https://doi.org/10.1037/xlm0000612>

Acknowledgements

This research began with a simple thought I had while walking along the Brooklyn Bridge on a crisp November New York morning - in the midst of the 2016 US Presidential election. I wondered, “Is it possible to get Republicans and Democrats to speak with each other in a civil manner?” Thus, began my quest for an answer – a treacherous journey into the depths of social psychology, which I embarked upon with such optimism. While I’m confident that I have found some answers, completing this thesis in the middle of a global pandemic, and enduring yet another US Presidential election in 2020, has indeed questioned the naïve optimism I had at the start of this journey. Maybe it’s a sunk-cost fallacy (maybe it’s not), but in hindsight, I do not regret a minute of my quest to understand the complexity of the human mind.

Of course, none of what I have discovered and accomplished in the last few years would be possible without a team of cheerleaders behind me. First and foremost, I would like to thank Dr. Simon Laham for being a wonderful supervisor and supporter. The last few years have certainly been challenging, but your guidance has been monumental in overcoming both professional and personal struggles. You continue to inspire me with your brilliant mind, your thirst for knowledge, and moral compass. Thank you for being the best supervisor, and for encouraging me to pursue my passions.

Thank you to the incredible faculty and my supervisory committee at the Melbourne School of Psychological Sciences. A special thank you to the PhD candidates that I met at the Melbourne Moral Psychology Lab, especially Dr. Damien Crone, Dr. Shaheed Azaad, and Joshua Rhee. You’re some of the smartest people I have met. I miss our academic and non-academic banter, and will cherish our friendship, despite being oceans apart. Thank you to my closest friends who have been by my side since undergraduate studies and all throughout this degree. A special thank you to Sanju Reddiar and Akla Nanayakkara, whom I was fortunate enough to share a home with - thank you for taking care of me during

tough times, and for pushing me to pursue my professional goals. A special thank you to Sumudu Mallawaarachchi for our entertaining writing sessions together, our fascinating philosophical discussions, and for always being there to wipe the many tears over the years. Thank you to Ashley Zahra and Jasmine Schipp for your loving and kind encouragement, and for cheering me up when I needed it the most.

A huge thank you to my family, especially to my parents, Las and Kris, and my sister, Dilini. You three continue to inspire me with your dedication to your respective fields. I wouldn't be where I am today without your support. You'll have always encouraged me to reach my fullest potential, and have instilled me with an insatiable appetite for learning and contributing to society to the best of my ability. I'm so grateful for the insane amount of time and energy you have put into me since my childhood.

Last but not least, thank you to my loving husband, Nishan. You stood by me the last three years, showering me with kindness during the most difficult times. While planning our wedding during the PhD was certainly challenging, marrying you was the best decision I have ever made. We are from completely different worlds, but your effort to understand what I do, help others understand what I do, and your ability to challenge my thinking and push through during bleak times has been astounding. None of the amazing things that I have accomplished during the last few years, including submitting this thesis, would have been possible without your strength and support. I believe that our relationship is a testament to the ability of people from contrasting backgrounds, different beliefs and socio-political attitudes, to compromise and make decisions when consensus is needed the most. Our ability to bridge our ideological divides successfully has given me the courage to dream of a utopian society.

Abstract

In the currently fractured attitudinal landscape, many people hold extreme attitudes on important socio-political issues and frequently advocate for their attitudes. As a result of advocating for one's own attitudes, i.e., engaging in pro-attitudinal advocacy, people may unconsciously persuade themselves (i.e. *self-persuade*) to polarize further. Moreover, pro-attitudinal advocacy may lead to an unwillingness to engage with those on the opposite side of the attitude spectrum. Such processes may explain polarization in our electorates over time.

Given that polarization leads to disrupted social cohesion, a critical question is how we might depolarize attitudes, and encourage open communication. In this thesis, I sought to examine a candidate mechanism of depolarization: self-persuasion. More specifically, this thesis explored self-persuasion occurring during pro-attitudinal advocacy as a way of a) depolarizing attitudes and b) opening up communicative practices in ways that encourage advocates to display open communicative intentions (e.g. willingness to engage with dissimilar others). Across seven studies, participants ($N = 1946$) were asked to engage in online advocacy to justify their attitudes towards contemporary socio-political and moral issues, such as migration and climate-change. I explored specific factors relevant to pro-attitudinal advocacy, including meta-cognitive confidence (confidence in one's thoughts and arguments), advocacy framing (moral versus practical framing), and style of advocacy (defending one's position versus attacking the opposing position).

Results indicate that a) low meta-cognitive confidence in one's advocacy attempt predicts attitude depolarization and greater receptiveness to opposing views, especially if one enjoys effortful cognitive activity (Studies 1 and 2), b) compared to moral justifications, using practical justifications to support one's attitude predicts attitude depolarization and lower intentions to proselytize, via lower expression of moral language (Studies 3-5), and c)

compared to defending one's position, attacking the opposing position predicts more open communicative practices, such as seeking attitude-inconsistent information, via heightened motivation to hold accurate worldviews (i.e. accuracy motivation; Studies 6 and 7).

Interestingly, our findings suggest that although attitudes sometimes depolarize, depolarization is not a necessary pre-requisite for open communicative intentions, such as seeking attitude-inconsistent information. This suggests that pro-attitudinal advocacy may influence behavior, even if attitudes remain unchanged. Further, in addition to self-persuasion, we find other mechanisms, such as, message content and accuracy motivation, which may predict depolarization and increase open communicative practices. We find evidence to suggest that re-framing advocacy attempts, such as encouraging practical as opposed to moral justifications, may depolarize attitudes and encourage more civil discourse between dissimilar others. Overall, our findings have implications for bridging gaps between opposing groups, particularly in online communication environments.

Chapter 1 Introduction

Society is extremely diverse in attitudes towards religious, socio-political, and moral issues, and people are apparently receptive to this diversity (Pew Research Center, 2008). However, people continue to express prejudice towards those holding different worldviews (Brandt, 2017). People with strong religious views, for example, tend to avoid others who hold different religious affiliations to their own (Hart et al., 2009). Similarly, people with strong political views often avoid or even degrade those holding different political ideologies – this is reflected by America becoming increasingly polarized over the last decade (Pew Research Center, 2014, 2016). Thus, despite apparent receptivity to diverse opinions, it is evident that people polarize away from those holding different opinions to their own.

This polarization is intensified by communicative practices, especially those online (Hestres, 2015), by providing accessible outlets for people to advocate for their own opinions (*pro-attitudinal advocacy*; Briñol, McCaslin, & Petty, 2012; Gordijn, Postmes, & de Vries, 2001). Advocating for one's own opinions, or *pro-attitudinal advocacy*, typically tends to polarize groups even further (group polarization; see Myers & Lamm, 1976; Sunstein, 2002 for reviews), resulting in “echo chambers” of like-minded individuals who reinforce each other's attitudes and beliefs (Sunstein, 2002). The formation of such political and cultural silos results in fragmented groups of people who are unwilling to engage with each other in a civil manner.

Thus, discourse today is fractured. Not only do people interact with like-minded others holding polarized attitudes, but they also tend to strongly advocate their attitudes. The 2016 US Presidential election and the 2017 same-sex marriage plebiscite debate in Australia were testaments to this observation. Such heated debates contribute to fragmentation of the electorate and disrupted social cohesion. In extreme cases, political polarization leads to violence, for example, on the part of both Republicans and Democrats during the 2016

presidential election (Araida et al., 2016). Politically-motivated violence is an obvious reflection of our fragmenting society, although more subtle prejudices also contribute to the divergence of opposing groups.

For example, the polarized political landscape appears to be reinforced by the use of moral rhetoric to justify one's attitudes. People often rely on arguments grounded in moral values more so than those grounded in economic consequences to substantiate their own position. This is to be expected, given that people think that their own moral values are both factual and universal (Tetlock, 2003), and that they have difficulty understanding worldviews which are different from their own (Feinberg & Willer, 2013, 2015). However, this is dangerous because using only moral reasoning as justification for one's attitudes tends to strongly motivate action, which manifests as persuading others, signing petitions, or in some cases, even hate crimes (Skitka & Mullen, 2002)

The fractured political landscape also appears to be reinforced by political leaders who use their media presence to cultivate resentment for opposing political candidates. Indeed, presidential nominees demonstrate a long history of using "attack politics" to garner support from voters by degrading their opponents (Haynes & Rhine, 1998). Possibly drawing on this behavior, nearly 50% of voters in the 2016 US presidential election deemed attacking and personally insulting the opposition as appropriate (Pew Research Center, 2016). In this way, advocacy often tends to be couched in negative terminology towards the opposition, as advocates tend to degrade the opposition to reinforce the legitimacy of their own position (Ellemers & van den Bos, 2012; Voelkel, Brandt, & Colombo, 2018), rather than elaborating on the merits of their own arguments.

Given the polarizing political landscape, it is evident that a solution is required to bridge gaps between opposing ideologies, by untying those who are entrenched in their own worldviews. One way of doing so, is by using persuasion to change attitudes and behaviors.

In this short introductory chapter, I will briefly review existing research relevant to depolarizing entrenched attitudes. First, I will review traditional methods of persuasion, i.e. “standard” persuasion, which have been used to change attitudes and behaviors. Then, I will review “emerging” methods of persuasion, i.e. self-persuasion, which I propose as a potential complement to standard persuasion methods. Next, I will elaborate on the particular self-persuasion paradigm central to the present research, i.e. pro-attitudinal advocacy, and highlight why this paradigm has the greatest potential to bridge ideological divides. Finally, I will end with the aims and findings of the present research. The primary goals of this thesis were to: (1) Understand the mechanisms through which pro-attitudinal advocacy leads to attitude depolarization, and increased willingness to engage with dissimilar others, and (2) Uncover ways in which pro-attitudinal advocacy may be re-framed in ways which depolarize attitudes and increase openness to the opposition.

A Review of “Standard” Persuasion

Persuasion plays a crucial role in educational and political reform, religion, and consumer behavior, and in these different contexts, people are often exposed to external persuasive messages (Petty & Briñol, 2008). Persuasion is ubiquitous and represents a civil means by which people’s attitudes can be changed (Petty & Cacioppo, 1981; 1986; Petty & Briñol, 2002; Petty & Briñol, 2008). In what I term “standard” persuasion, a recipient (person or group of people) is subject to a persuasive message from an external source. That is, the recipient of the persuasive message and source are two different entities. If the persuasive attempt is successful, then the recipient’s attitude will change. For example, a political conservative receiving a message from a conservative party leader may become more

polarized in their opinion, if the persuasive message delivered contains strong arguments which aligns with their values.

The most prominent theories of standard persuasion which explicate this attitude change, are *dual-process* theories, of which the Elaboration Likelihood Model (ELM) has received the most scholarly attention in the persuasion literature (Petty & Cacioppo, 1984a; Petty & Cacioppo, 1986; see Petty & Briñol, 2014 for a recent review). This model suggests two “routes” to persuasion – a central and peripheral route, by distinguishing between recipients who are able and motivated to process a persuasive message, compared to those who are unable to do so. Whether a recipient is able or motivated to process a persuasive message depends on factors such as, the personal relevance of the message, and individual predispositions, such as, Need-for-Cognition (NfC; those high in NfC enjoy engaging in effortful cognitive activity; Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984; Cacioppo, Petty, Kao, & Rodriguez, 1986; Cacioppo, Petty, & Morris, 1983). In other words, the ELM suggests that attitude change processes and factors vary depending on the qualities of the recipient.

If the recipient is motivated and able to process a persuasive message, he/she is said to follow the central route to persuasion, whereby attitude change is affected by the quality of arguments presented, and the credibility of the source delivering the message, for example, if the recipient is unmotivated or unable to pay attention, then attitude change is affected by heuristic cues, such as source attractiveness (Cacioppo & Petty, 1989; Petty & Cacioppo, 1984b; Petty, Cacioppo, & Schumann, 1983). In addition to kinds of variables (e.g. arguments quality, source credibility) which may influence attitude change, the ELM also outlines the processes through which attitudes change under each of the two processing routes. For example, attitude change is driven by peripheral, heuristic cues during peripheral route processing, such as, source attractiveness (Petty & Cacioppo, 1984a). On the other

hand, attitude change is driven by more thoughtful, deliberative processes during central route processing, such as, the amount of confidence people place in their thoughts following exposure to the persuasive message (see thought confidence; Briñol & Petty, 2003; Petty, Briñol, & Tormala, 2002; Briñol, Petty, & Tormala, 2004; Tormala, Briñol, & Petty, 2007; Tormala, Briñol, & Petty, 2006). If people are confident in the thoughts that they generate during advocacy, they are more likely to polarize, while if they are not confident in their thoughts, they are more likely to depolarize.

Taken together, work in standard persuasion shows that changing aspects of the persuasive message or source to match characteristics of the audience (either their motivation, ability, or individual predispositions) will change attitudes in the direction intended by the source. Further, the ELM is useful because a) it provides a blueprint for which kinds of variables are most useful in convincing particular audiences and b) details the processes through which attitude change occurs depending on qualities of the audience (e.g. heuristic processes versus thought confidence). The ELM has been used to change attitudes towards a broad range of issues, including consumer products, university policies, and recycling, and has received attention in multiple fields outside of social psychology, such as in the communications literature and also in marketing and management research (Briñol et al., 2004; Haugtvedt & Petty, 1992; Petty, Cacioppo, & Goldman, 1981; Tormala et al., 2006). The ELM will be examined in greater detail in Chapter 2.

A Review of Self-Persuasion

An alternative to standard persuasion as conceptualized by dual-process theories, such as, the ELM, is *self-persuasion*. I examine self-persuasion as an important alternative to changing attitudes and behaviors, because as highlighted earlier in this chapter: a) society is becoming increasingly polarized in their attitudes towards socio-political views (Pew Research Center, 2014), and partisans tend to hold attitudes that are highly resistant to

change, and b) attitudes that are morally entrenched and held with high confidence are more resistant to change using standard persuasion (Krosnick & Petty, 1995). Thus, I consider self-persuasion as a more subtle method of driving attitude and behavior change.

The distinction between standard and self-persuasion is that in the latter, the recipient and source are the *same* person or group of people. This implies that the message is both generated and received by one entity. In other words, the target of persuasion is the same person who is generating the persuasive message. This is an interesting distinction because it implies that the recipient has a large degree of control over the persuasive process and one's own attitude change. On the flipside, it also implies that the experimenter has a degree of control over the circumstances in which a person generates the message, and how he or she does so, and thus some control over the outcomes of persuasion: attitude or behavior change. To study self-persuasion, many experimental paradigms have been adopted to date – I will briefly review each of these (see Chapter 3 for a more in-depth investigation of these paradigms).

Group Polarization

In the group polarization paradigm, people generate persuasive messages in a group (see Isenberg, 1986; Myers & Lamm, 1976; Sunstein, 2002 for reviews). The key tenet of group polarization is that people's attitudes will shift to resemble the group consensus. That is, people's privately held attitudes will change as a result of being exposed to other group members' attitudes, and that attitudes will polarize in the direction implicated by the group. Overall, the group polarization literature points to a *risky shift* whereby attitudes tend to shift in the direction of group norms, which typically tend to be "riskier" or more polarized than initially held attitudes (Moscovici & Zavalloni, 1969; Stoner & Finch, 1961; Stoner, 1968). What this means in the current political context, is that a group member who is initially slightly pro-Trump becomes extremely pro-Trump after speaking with fellow Trump

supporters. The problem with group polarization is that it tends to produce echo chambers of those holding similar views (Sunstein, 2002) which can be dangerous, as echo chambers impede progress requiring collaboration between those holding different views. One way of combatting problems created by group polarization is by introducing a third party to the group to moderate the conversation (Standberg, Himmelroos, & Gronlund, 2017). A more in-depth investigation on how group polarization can be leveraged to depolarize attitudes can be found in Chapter 3.

Mere Thought

Another experimental paradigm investigating self-persuasion is mere thought – the mere thought effect occurs when people are simply asked to think about an attitude issue (not verbalize orally or in writing; see Tesser, 1978 for a review). This line of research shows that the longer one thinks about the attitude issue, the more likely one is to polarize. For example, 35% of participants polarize after thinking about an attitude issue such as “prostitution should be legalized” for 30 seconds, but this number jumps to 60% when participants are allowed to think for 180 seconds, due to the generation of more attitude-consistent thoughts (Tesser & Conlee, 1975). Interestingly, recent research shows that depolarization can be induced in mere thought paradigms when participants are allowed to think for even longer periods of time (e.g. 300 seconds), possibly because participants then begin to question their own attitude-consistent thoughts (i.e. express low thought confidence; Clarkson, Valente, Leone, & Tormala, 2013).

Saying-is-Believing

Work in the communications literature provides evidence for *audience design*, i.e. that speakers are biased as they invariably take into account their audience’s attitudes when designing their messages (Bell, 1984; Clark & Murphy, 1982). This bias is consequential because designing a communication in-line with the audience affects the attitudes of the

speaker themselves (i.e. a “saying-is-believing” effect). For instance, speakers who are led to believe that their audience likes the attitude object that they are describing, tend to design their message to contain more positively valenced words (Fussell & Krauss, 1989b, 1989a). This biased message description is then consequential. For example, when a speaker is asked to describe a target to a colleague to enable him/her to choose a roommate, if the speaker uses more positively valenced words to describe the target, they are more likely to display favorable attitudes toward the target themselves (Higgins, 1992; Higgins & Rholes, 1978). This means that even simply describing a target to an audience can affect one’s attitudes towards the target (see Chapter 3 for exceptions).

Self-Reflection

Another line of work in self-persuasion shows that people’s attitudes change when they are asked to reflect on reasons for their attitudes towards people or objects (see Wilson, 1990 for a review). Interestingly, this attitude change tends to be in the form of depolarization, because people often find it difficult to explicate reasons for their own attitudes (Nisbett & Wilson, 1977). Attitude depolarization via self-reflection is additionally moderated by people’s knowledgeability on the issue – for example, when voters are asked to reflect on why they like their preferred political candidate, those who possess limited knowledge on the candidate are more likely to depolarize (Wilson, Kraft, & Dunn, 1989).

Counter-Attitudinal Advocacy

In this paradigm, advocates are forced to generate arguments in favor of the *opposing* position to their own, which typically results in attitude depolarization (Hovland, Janis, & Kelley, 1953). The idea here is that when people engage in counter-attitudinal advocacy, there is a mismatch between their original attitude and their behavior, resulting in an uncomfortable mental state of cognitive dissonance (Harmon-Jones & Mills, 1999). Thus, people are compelled to change their attitude in the direction implied by their arguments, in

order to restore consonance between their attitude and behavior (Aronson, 1969; Festinger & Carlsmith, 1959). The effect of cognitive dissonance on attitude depolarization tends to be amplified if people generate their own arguments from scratch (i.e. improvise their arguments, as opposed to regurgitate experimenter-provided arguments), or when people are less knowledgeable about the topic at hand (e.g. unfamiliar issue: duration of the common cold versus familiar issue: meat consumption in the US; Janis & King, 1954; King & Janis, 1956). Although a recent review shows that counter-attitudinal advocacy tends to be an effective method of depolarizing attitudes (Kim, Allen, Preiss, & Peterson, 2014), this form of advocacy is not ecologically valid: people do not often spontaneously advocate for an opponent's position on an issue, rather, they tend to advocate for their own position on the issue, i.e. *pro*-attitudinal advocacy.

We propose self-persuasion as a potentially more effective alternative to standard persuasion in the current political climate, given that attitudes towards highly moralized issues, such as, migration, and climate-change, tend to be held with high confidence. Different persuasion paradigms can be used to depolarize attitudes and increase openness to opposing viewpoints, including, group polarization, mere thought, saying-is-believing, self-reflection, and counter-attitudinal advocacy. However, we contend that self-persuasion induced by *pro*-attitudinal advocacy provides a more subtle avenue to induce depolarization, given that people often spontaneously advocate for their own views.

Pro-Attitudinal Advocacy and Self-Persuasion

This thesis specifically focuses on a single self-persuasion paradigm: *pro*-attitudinal advocacy (PAA). While in counter-attitudinal advocacy paradigms, people advocate for the opposing position to their own, in *pro*-attitudinal advocacy paradigms, people advocate for their *own* position on the topic. The reason we focus on this type of advocacy is because PAA is extremely understudied compared to other types of advocacy, and yet represents the most

ecologically valid type of advocacy in today's society. People advocate for their position frequently, be it face-to-face (e.g. parliamentary debates) or online (e.g. Reddit, and other social media platforms). This presents us with an opportunity to first, understand what happens when people advocate for their opinions, and second, to explore how to re-frame pro-attitudinal advocacy in ways which encourage people to question their own attitudes. These two questions are central to the key motivation behind this thesis: how can we encourage greater cooperation between opposing ideological groups?

The limited existing literature in PAA shows that people typically polarize when advocating for their own opinion, but this can be attenuated by manipulating the context in which the advocacy occurs. For instance, informing people that their attitudinal position is a minority position as opposed to a majority position leads to attitude depolarization, due to social validation concerns (i.e. low consensus must mean incorrectness; Gordijn et al., 2001).

In addition to social consensus, changing the target of advocacy has been shown to influence depolarization. Specifically, asking people to aim the advocacy attempt at oneself as opposed to a hypothetical target audience results in greater attitude depolarization (Briñol et al., 2012). This is because people are less accustomed to persuading themselves, therefore put less effort into their appeal and thus perceive the quality of their message to be weaker. In other words, people may experience low confidence in their advocacy attempt, i.e. low meta-cognitive confidence, which is the amount of confidence people place in their thoughts, and other attitude-relevant judgements, such as, confidence in their arguments (see Petty, Briñol, Tormala, & Wegener, 2007; Wagner, Briñol, & Petty, 2012 for reviews). This suggests that undermining the confidence that people have in their advocacy attempts by using false feedback mechanisms, for example (see Petty, Briñol, & Tormala, 2002; Tormala, Briñol, & Petty, 2004), may promote depolarization (see Chapter 3 for a more in-depth discussion of extant work in PAA).

Meta-Cognitive Confidence, Elaboration, and Self-Persuasion.

Taken together, extant work in PAA shows that attitude depolarization may be attenuated when some adjustments are made to the advocacy context. However, there are a few key limitations to extant research: a) other than for the effects of numerical support and target, there is currently low knowledge of specific factors that may induce depolarization during PAA. That is, we do not know the ways in which PAA appeals might be reframed in ways that lead to depolarization, b) we also do not know the mechanisms by which attitudes change during PAA, other than for meta-cognitive confidence, and c) we do not know the consequences other than for attitude depolarization, such as behavioral intentions, which may be influenced by PAA. Examining behavioral intentions (hereafter referred to as communicative intentions) is important because, even though it is difficult to change attitudes which are held with high conviction (Krosnick & Petty, 1995), communicative intentions may nevertheless change as a result of advocacy. For example, previous work examining the effects of legalizing same-sex marriage in the US, has shown norm perceptions and subsequent behavior towards gay people to change, even when attitudes do not change (Tankard & Paluck, 2017).

Thus, in addition to attitude depolarization, we examine the consequences of pro-attitudinal advocacy on a suite of communicative intentions, including positive, “open” communicative intentions, such as, choosing interaction partners with opposing views (Rios, Goldberg, & Totton, 2018), and seeking attitude-inconsistent information (Hart et al., 2009), and also negative, “closed” communicative intentions, such as, intentions to proselytize one’s opinions (Cheatham & Tormala, 2015, 2017).

To this end, we start by examining whether meta-cognitive confidence predicts attitude depolarization and communicative intentions following PAA on a controversial attitude issue related to climate-change (see Chapter 4 for details). The key difference

between our work and previous work exploring meta-cognitive confidence and PAA (see Briñol et al., 2012), is that we examine attitude (de)polarization and communicative intentions (e.g. intentions to engage with those holding opposing views) in a pre-post design, as opposed to a between-subjects design, and on a controversial attitude issue (as opposed to a less polarized attitude issue, such as university policies). Specifically, we predict that when meta-cognitive confidence in one's advocacy attempt is low, one is more likely to depolarize post-advocacy. In addition, we examine what other factors related to confidence and elaboration are likely to predict depolarization and communicative intentions. Specifically, we examine the effects of Need-for-Cognition (NfC; individual tendency to enjoy effortful cognitive activity, see Cacioppo & Petty, 1982; Cacioppo et al., 1984 for reviews), which, to our knowledge, is yet to be explored in the context of self-persuasion.

High NfC individuals who enjoy cognitive activity, are more likely to display increased meta-cognitive confidence in their advocacy attempts due to their increased tendency to think about complex issues (Briñol, Loersch, & McCaslin, 2009). On the flipside, they are also more amenable to revise their beliefs when their confidence is low (Wegener & Petty, 1997). Thus, we predict that people who are high in NfC are more likely to depolarize and express greater receptivity to opposing views when confidence in their own advocacy appeal is low. We explore this possibility in Chapter 4.

Manipulations of PAA

Given the lack of work in social psychology investigating PAA, to form predictions of the types of experimental manipulations that might be worth examining, we turn to work in political psychology and communications research.

Moral vs Practical. Political psychology suggests that a dimension along which PAA can vary is content of advocacy. Specifically, advocacy appeals may vary in the types of arguments used as justification for one's position, such as moral versus practical arguments.

Although no work in self-persuasion has directly examined the consequences of moral versus practical advocacy, work intersecting political psychology and standard persuasion suggests that advocacy couched in morality as opposed to practicality likely has a stronger impact on attitude change.

For instance, Leidner, Kardos, and Castano (2018) have shown that presenting moral arguments (arguments grounded in rules, values, moral principles) compared to practical arguments (arguments cost-benefit analyses, economic consequences), is more likely to change attitudes towards controversial issues such as, capital punishment. This suggests that arguments grounded in morality are more powerful in influencing one's attitudes, compared to those grounded in practical reasons.

This is because moral values are ingrained, and inform one's worldviews, making moral rhetoric extremely persuasive and divisive (Kovacheff, Schwartz, Inbar, & Feinberg, 2018; Tetlock, 2003). Moral rhetoric tends to be persuasive given its ability to evoke strong emotions such as, anger and disgust (Haidt, 2003), its ability to connect with people's existing moral values (Feinberg & Willer, 2015; Voelkel & Feinberg, 2018), and its ability to appeal to a variety of moral systems or frameworks such as, rules, and principles, but also "moral" outcomes and consequences (Wheeler & Laham, 2016). Thus, morally framed messages tend to result in polarization and also proselytization of one's opinions (Clifford, 2019; Ryan, 2017; Tappin & McKay, 2019; Valenzuela, Piña, & Ramírez, 2017).

Despite understanding the detrimental consequences of moral advocacy, little is known as to a) *how* moral advocacy leads to polarization and proselytization (i.e. mechanisms), b) whether moral advocacy leads to such consequences as a result of self-generated persuasion, and c) how advocacy can be re-framed in ways which encourage free speech but minimize social conflict. Thus, we seek to establish the mechanisms through which moral advocacy leads to social division, while exploring an alternative method of

advocating for one's attitudes. To this end, we test the mechanisms and effects of practical versus moral advocacy using text-based analysis in Chapter 5.

Attack vs Defend. In addition to moral versus practical advocacy, a separate line of work in the communications literature suggests that there are additional variations of PAA, and that these variations have different consequences on people's attitudes and communicative intentions. To elaborate, communications research distinguishes between "advocacy" and "refutation" (Infante & Rancer, 1993; Kuhn, 1991). To illustrate the difference, consider a debate - you may either justify your own stance, or refute your opponent's stance. Although work in this space has focused on the antecedents of advocacy versus refutation practices (e.g. people who are generally more argumentative are more likely to use both advocacy and refutation to justify their position; see Rancer & Avtgis, 2006 for a review), this work fails to examine the downstream consequences of these two types of advocacy on the speakers themselves.

This distinction between justifying one's own position versus attacking someone else's position has also been drawn out in political advertising theory, which shows that political candidates can either engage in positive advertising "you should vote for me because..." or negative advertising "you shouldn't vote for my opponent because..." (see Moorman & Neijens, 2012 for a review). Again, this literature fails to examine the effects of positive or negative advertising on the speaker themselves. Given that speaker-generated messages have been shown to impact the speaker (as shown by early work in self-persuasion), we speculate that distinguishing between defending oneself (we term this Defense) versus attacking the opposition (we term this Attack) in pursuit of justifying one's own position, should have different effects on oneself as well.

For example, it is possible that Attacking one's opponent's position will lead to greater feelings of anger, attitude polarization, and thus greater willingness to proselytize

one's opinions. In contrast, it is possible that defending one's position will lead to low meta-cognitive confidence, given that people have difficulty explicating reasons for their own opinions (see work on self-reflection above; Wilson, 1990). This low meta-cognitive confidence may then translate into depolarization, and greater willingness to interact with those holding opposing views. We will examine various manifestations of Attack vs Defense, including the consequences and mechanisms of Attack versus Defense in the current research (see Chapter 6).

The Current Research

This thesis examines the possibility of weakening polarized attitudes (i.e. depolarizing attitudes) in ways which allow people to become more open to opposing viewpoints. Given that it is difficult to change highly convicted attitudes (Krosnick & Petty, 1995), I also seek to explore ways of encouraging more open communicative practices among advocates, even if their attitudes do not depolarize, such as, seeking attitude-inconsistent information, expressing greater willingness to interact with the opposition, and lowering intentions to proselytize one's opinions. Thus, the two key aims of the current research are as follows: (1) Understand the mechanisms through which attitude depolarization and open communicative practices results following a pro-attitudinal advocacy attempt, and (2) Uncover ways to re-frame pro-attitudinal advocacy attempts in ways which depolarize attitudes, increase open practices, such as, seeking attitude-inconsistent information, and decrease closed practices such as, proselytization.

In light of Aim 1, I consider multiple mechanisms through which one's communicative practices may change, including, self-persuasion (see Chapter 4), content of one's speech (see Chapter 5), and motivations (see Chapter 6). In light of Aim 2, I explore two novel ways of re-framing pro-attitudinal advocacy attempts in ways which encourage free speech but simultaneously have the potential to reduce social conflict, including moral

versus practical advocacy (see Chapter 5), and Attack versus Defense advocacy styles (see Chapter 6).

This thesis is structured as follows: In Chapter 2, I begin by exploring the history of attitudes and attitude change, and examining more closely the standard persuasion methods which have traditionally been used to change attitudes. In Chapter 3, I then explore an alternative to standard persuasion, which is self-persuasion, and briefly review the literature in this area. I also explore in more detail, the specific self-persuasion paradigm most relevant to the current research: pro-attitudinal advocacy.

The next three Chapters to follow (4, 5, 6) consist the primary empirical chapters of this research (inserted as accepted or submitted publications). In Chapter 4, I conduct an empirical test of self-persuasion in the context of pro-attitudinal advocacy, and find support for our prediction that low meta-cognitive confidence leads to depolarization post-advocacy. We also establish, for the first time, that Need-for-Cognition moderates this effect, such that those high in NfC are more likely to depolarize, and express willingness to engage with those holding opposing views, when their confidence is low. In Chapter 5, I conduct the first empirical test of moral versus practical advocacy in the context of pro-attitudinal advocacy, and as predicted, find that practical advocacy leads to greater depolarization, and lower intentions to proselytize one's opinions, via lower levels of moral expressiveness. In Chapter 6, I conduct the first empirical test of Attack versus Defense advocacy styles in the context of pro-attitudinal advocacy. Interestingly, we find that those who employ Attack advocacy are more likely to overcome confirmation bias, and display increased receptivity to the opposition, compared to Defense advocates. This effect is explained by accuracy motivation, such that, those who attack are induced with a greater motivation to hold accurate worldviews following their advocacy attempt, compared to defense advocates. Finally, I end on Chapter

7, with a general discussion on how we can harness the power of pro-attitudinal advocacy to potentially bridge ideological divides.

Chapter 2 History of Attitudes & Attitude Change

The study of attitudes and attitude change has been and continues to be, central to social psychology. Given the decades of research dedicated to understanding attitude change, it is unsurprising that the conceptualization of attitudes has evolved markedly over time. One of the first definitions of attitudes, articulated by Allport (1935) states that: “an attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual’s response to all objects and situations with which it is related” (p.784). Extending Allport’s (1935) broad definition, in later years, attitudes were referred to as cognitions, emotions, or behavioral reactions generated towards objects or situations (see Zanna & Rempel, 1988).

However, since the cognitive revolution, the definition of attitudes has narrowed substantially; subcomponents of attitudes such as affect and behavior were stripped out of the definition, and attitudes are now simply defined as a summary evaluation of an attitude object (Petty & Cacioppo, 1981, 1986; Eagly & Chaiken, 1993; Fazio, 1995). In other words, an attitude is a “psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly & Chaiken, 1993). That is, an attitude refers to a general favorable or unfavorable predisposition towards an attitude object (i.e. another person, a group of people, inanimate being, a policy, a lifestyle, occurrence of an event, and so forth). This definition represents the most contemporary articulation of attitudes in the current study of social psychology (Albarracin, Johnson, & Zanna, 2014), and the definition that is most closely linked with the study of persuasion and attitude change.

Commensurate with the contemporary definition of attitudes, attitude change in the current context refers to a shift in the evaluation of a particular object, following a persuasive intervention (Petty & Briñol, 2015). A typical persuasive attempt comprises a person or a group of people (recipients) who experience an intervention (persuasive message) from

another person (message source) with regards to a particular attitude object (Petty, Priester, & Briñol, 2002). A persuasive message typically contains positive or negative arguments related to the particular attitude object. At the time of receiving the persuasive message, a person may have a particular evaluation of this attitude object, but as a result of exposure to the message, this evaluation may change. That is, attitude change occurs as an outcome of people experiencing persuasive messages. A persuasive attempt is deemed to be successful once the recipient's attitude changes in the direction intended by the message source (Petty et al., 2002).

To summarize, attitude change can be achieved by altering the link between an object and its evaluation via persuasive processes (see Figure 2-1)¹. The next question is, what factors and processes are responsible for attitude change?

¹ For the purpose of this thesis, we will be using the terms 'attitude change' and 'persuasion' interchangeably.

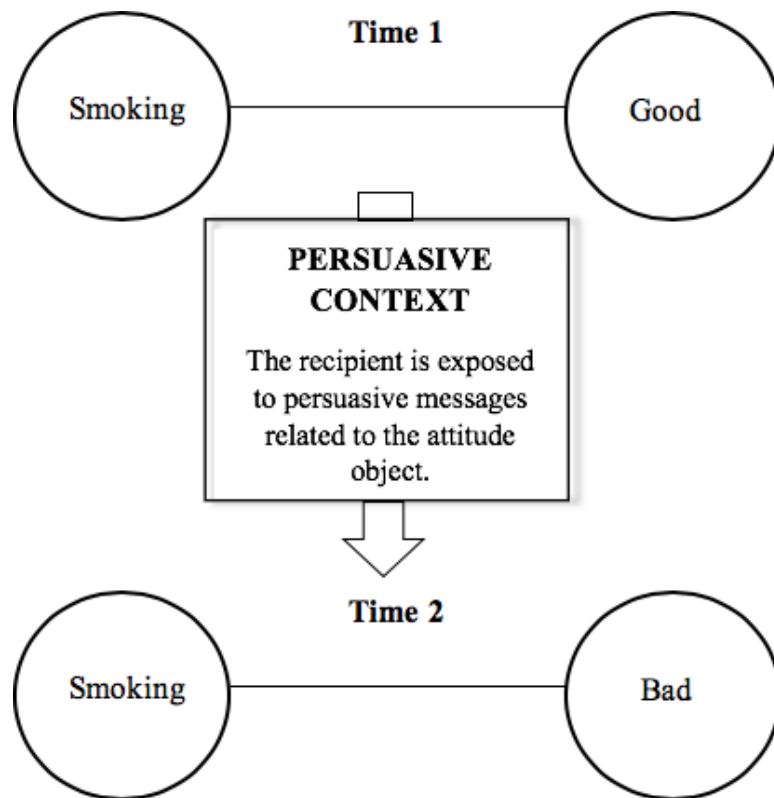


Figure 2-1. A schematic representation of attitude change.

At Time 1 the recipient perceives smoking (the attitude object) as positive (evaluation). By exposure to a persuasive intervention, at Time 2 the recipient perceives smoking as negative (example adapted from Petty, Briñol, & DeMarre, 2007).

Traditional Theories of Persuasion

A brief look at history points to an interesting birth of persuasion research. The first systematic approach to identifying factors responsible for attitude change was taken by Carl Hovland and his colleagues through the Yale Communication Program (YCP). This research program was primarily established to understand the effectiveness of political propaganda on soldiers (McGuire, 1968), although research was conducted on attitudes towards medicine,

meat consumption, consumer products, and so forth (see Hovland, Janis, & Kelley, 1953 for a review). Following a series of experiments, the researchers at YCP identified three classes of factors responsible for attitude change: message source (e.g. characteristics of the person delivering the message, such as, credibility), the message (e.g. the strength and number of arguments contained in the message), and finally the recipient (e.g. characteristics of the audience, such as, individual traits and dispositions; see Hovland et al., 1953; Hovland, Lumsdaine, & Sheffield, 1949; Hovland & Weiss, 1953; Kelman & Hovland, 1953). According to YCP research, attitude change occurs as a result of manipulating one or more of these factors (see Figure 2-2). Put simply, these factors represent ‘*who says what to whom and with what effect*’ in persuasive contexts (Smith, Lasswell, & Casey, 1946).

Following decades of research in attitude change, these three factors have remained central to newer theories of persuasion (see dual-process models of persuasion; Chaiken & Trope, 1999; Chaiken, Liberman, & Eagly, 1989; Petty & Wegener, 1999; Petty & Cacioppo, 1981, 1986). Interestingly, manipulating one or more of these factors sometimes causes recipients’ attitudes to become more extreme, i.e., polarize in the direction of the recipient’s initial attitude. For example, highly credible or “expert” message sources tend to polarize attitudes (Briñol et al., 2004; Chaiken & Maheswaran, 1994; Heesacker, Petty, & Cacioppo, 1983; Kelman & Hovland, 1953; Petty, Cacioppo, & Goldman, 1981), as do persuasive messages containing high quality arguments (Haugtvedt & Petty, 1992; Petty, Haugtvedt, Heesacker, & Cacioppo, 1995; Petty & Cacioppo, 1986). Attitudes similarly polarize when recipients are in a happy compared to a neutral or sad mood during message processing (Briñol, Petty, & Barden, 2007; Petty, Schumann, Richman, & Strathman, 1993; Wegener, Petty, & Klein, 1994).

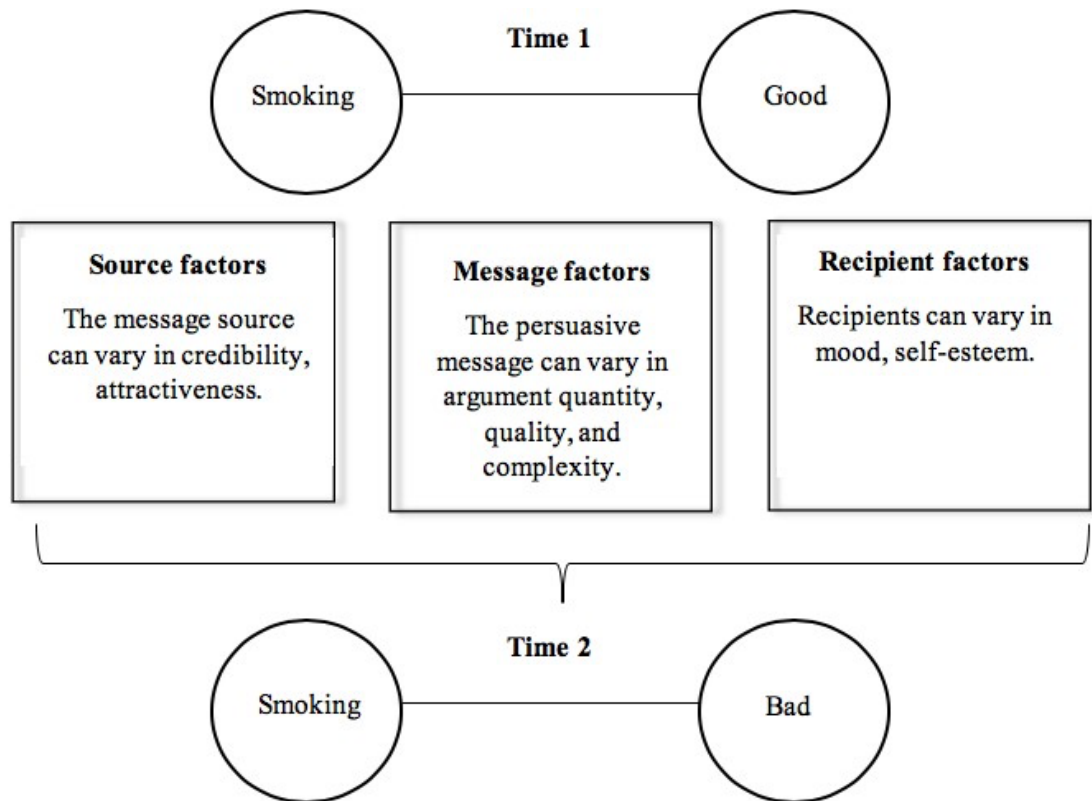


Figure 2-2. A schematic representation of factors affecting attitude change. At Time 1 the recipient perceives smoking (the attitude object) as positive (evaluation). At Time 2 the recipient perceives smoking as negative, influenced by characteristics of the persuasive source, message and/or the recipient's traits and predispositions.

Even more interesting, recipients' attitudes can sometimes change in the opposite direction to their initial attitude (i.e. depolarize) under certain circumstances. For example, recipients tend to depolarize when the message-presenting source is low in credibility (i.e. not a subject-matter expert, or not trustworthy (Briñol et al., 2004; Priester & Petty, 1995; Tormala, Briñol, & Petty, 2007), when the persuasive message contains low quality arguments (Petty, Haugtvedt, Heesacker, & Cacioppo, 1995; Petty et al., 1981; Tormala,

Briñol, & Petty, 2006), or a low number of arguments (Petty & Briñol, 2008; Petty & Cacioppo, 1984b)². This suggests that depolarizing attitudes via persuasion may lead people to become more open to opposing viewpoints, by moving closer to the opposition on the attitude spectrum.

Attitude change processes. The next question is, *how* do different persuasion variables (e.g. argument quality, argument quantity) subsumed under each of the three factors (i.e. message, source, recipient) drive attitude change? That is, what mediating processes operate to produce attitude (de)polarization? Cognitive response theory (CRT) contends that attitude change occurs as a result of recipient-generated thoughts towards the persuasive message (Greenwald, 1968; Petty, Ostrom, & Brock, 1981). According to CRT, recipients generate thoughts in response to the persuasive message, which are then integrated into their existing body of cognitions, attitudes, feelings and so on. The valence of the cognitive response is determined by the content of the persuasive message, its delivering source, and characteristics of the recipient (Greenwald, 1968). The key tenet of CRT is that, for attitude change to occur, recipients must actively rehearse “attitude-relevant cognitions” at moments during which the attitude object is made salient in the persuasive message (Greenwald, 1968). A persuasive message that evokes primarily favorable thoughts towards an attitude object will result in attitude polarization, while a message that elicits largely unfavorable thoughts will result in attitude depolarization.

Accordingly, CRT implies that recipients actively deliberate over the central features of a persuasive message, such as, message content. However, cues peripheral to message content, such as, characteristics of the message source (e.g. source attractiveness) may influence attitude change as well (Todorov, Chaiken, & Henderson, 2002). In the latter cases,

² A caveat to these findings is that different factors may sometimes interact to produce attitude change in the opposite direction to traditional expectations. For example, high credibility sources sometimes induce attitude depolarization if the arguments presented are weak (see Tormala et al., 2006). For simplicity’s sake we strip these interactions for now, but explicate these interaction effects later in this chapter.

recipients are processing the persuasive message superficially (Todorov et al., 2002) as opposed to actively deliberating over message content. Thus, a shortcoming of the CRT is that it presumes that all persuasion contexts involve a recipient who is actively engaged with message processing. This theory does not account for situations in which persuasion occurs in the absence of active deliberation.

Dual process models of persuasion. Unlike CRT, dual-process models of persuasion which emerged in the 1980's demonstrate that attitude change occurs even in the absence of active deliberation (see Heuristic-Systematic Model; Chaiken & Trope, 1999; Chaiken, 1980; Chen, Duckworth, & Chaiken, 1999; see Elaboration Likelihood Model of Persuasion; Petty & Cacioppo, 1981, 1983, 1986). The major merits of dual-process models are that they clearly illustrate the processes or mechanisms underlying attitude change, and that they provide clear evidence for attitude change occurring when recipients are both actively or passively involved in message processing. That is, in contrast to early theories which suggested that people need to be engaged in active, reflective thinking for attitude change to occur (see cognitive response theory; Greenwald, 1968; Love & Greenwald, 1978; McGuire, 1968), dual-process theories posit that people's attitudes may change even without people actively generating thoughts in response to persuasive message content.

According to dual-process models, recipients in a persuasive context typically engage in one of two "routes" or "streams" of cognitive processing; a central route (systematic stream) or a peripheral route (heuristic stream). Recipients engaged in systematic processing expend high levels of cognitive effort to assess the validity of a persuasive message and the message's conclusion (Chaiken, 1980; Chaiken et al., 1989; Chen & Chaiken, 1999). That is, recipients actively deliberate on message content when engaged in systematic processing. On the other hand, recipients engaged in heuristic processing expend minimal cognitive effort to assess the validity of a persuasive message. Instead, they rely on peripheral cues (such as

source attractiveness, likeability or expertise) in making judgements of the attitude object (Eagly, Wood, & Chaiken, 1978; Chaiken, 1987).

Central and peripheral routes of processing therefore vary in terms of expendable cognitive effort in driving attitude change. This is akin to other dual process models, such as, System 1 vs System 2 thinking in cognitive psychology (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974; Kahneman, 2011). In the ‘Systems Thinking’ approach to decision-making, people make judgements based on heuristic cues (e.g. “consensus implies correctness”) when following System 1 thinking. On the other hand, people make decisions based on active deliberation and reasoning processes when following System 2 thinking. In the context of persuasion, people following a heuristic processing stream may rely on heuristics such as, “experts can be trusted” in deducing the validity of a persuasive message, as opposed to actively scrutinizing message content (Chaiken, 1980; Chaiken, 1987; Chaiken, Liberman, & Eagly, 1989).

Whether a recipient follows the central (systematic) or peripheral (heuristic) processing route depends on level of elaboration, which is the amount of thinking recipient’s do with regards to the attitude object (see Petty & Cacioppo, 1986; Briñol & Petty, 2015 for reviews). Level of elaboration is driven by whether the recipient is motivated and/or able to process the persuasive message. Recipients are more motivated to process a persuasive message and thus take the central route to processing when the message is personally relevant, for example (Petty & Cacioppo, 1979a, 1986, 1990). Similarly, recipients are more able to process a persuasive message when they possess personality characteristics such as, Need-for-Cognition, which predisposes them to enjoy effortful cognitive processing (Cacioppo, Petty, & Kao, 1984; Cacioppo, Petty, Kao, & Rodriguez, 1986; Cacioppo, Petty, Feinstein, & Jarvis, 1996). Thus, changing the level of elaboration by manipulating

motivation and ability influences whether recipients are more or less likely to follow the central or peripheral route to persuasion.

The tripartite persuasion factors identified previously (i.e. message source, message, and recipient) polarize attitudes by serving different roles depending on the route to persuasion. For example, source attractiveness typically serves as a peripheral cue when recipients are low in elaboration (triggering the peripheral route), however, source attractiveness may be processed under high elaboration (triggering the central route) by serving as an argument for the position being advocated (Petty & Cacioppo, 1984a). This is especially true if the source is related to the persuasive message content. For example, an attractive source may act as a persuasive argument for the advocacy of a beauty product, by serving as a “visual testimony” for its effectiveness (Petty & Cacioppo, 1980 as cited in Petty & Cacioppo, 1984).

Similar to source attractiveness, argument quantity acts as a peripheral cue when recipients are unable or unmotivated to process information (Petty & Cacioppo, 1984b). That is, under low elaboration, people may merely count the number of arguments and agree with the position being advocated when more arguments are presented, regardless of their merit (Petty & Briñol, 2008). On the other hand, individuals who are more motivated and able to think (those under high elaboration) focus on argument quality and cogency. Therefore, adding weak arguments to a message increases persuasion under low elaboration, however, decreases persuasion under high elaboration conditions (Alba & Marmorstein, 1987; Petty & Cacioppo, 1984b). This suggests that different factors (i.e. message properties and recipient characteristics) may interact when producing attitude change. Therefore, it is important to a) identify which combinations of variables are most likely to lead to attitude depolarization, and b) investigate any previously unconsidered variables (e.g. specific message content) which may drive attitude depolarization. We elaborate on these ideas later in this chapter.

Mechanisms for attitude depolarization. Thus far we have explored: a) the classes of factors (message source, message, recipient) which drive attitude depolarization, and b) routes to message processing (central and peripheral route) which recipients take when processing a persuasive message. Now we consider how specific persuasion variables under each of the tripartite factors drive attitude depolarization when recipients are following each of the two processing routes.

The ELM and HSM identify multiple roles for a single persuasion variable (Petty & Briñol, 2012; Todorov, Chaiken, & Henderson, 2002). For example, when level of elaboration is low, source credibility may serve as a simple heuristic cue (via the peripheral route), but when elaboration is high, source credibility may: a) affect the amount of information processing, b) bias the valence of thoughts generated towards the attitude object, c) serve as a piece of evidence for an argument, or d) influence confidence in thoughts generated (see Briñol et al., 2004; Heesacker et al., 1983; Petty et al., 1981; Petty & Briñol, 2002; Priester & Petty, 1995; Tormala et al., 2007; Tormala et al., 2006).

Recipients under low elaboration and thus following the peripheral route to persuasion, typically depolarize by using message sources and persuasive messages as simple heuristics. These heuristic cues include source credibility (expert or trustworthy sources), source likeability or attractiveness (popularity affecting agreeableness with message content), and length of the persuasive message (more arguments implying greater validity; see Axsom, Yates, & Chaiken, 1987; Chaiken, 1980, 1987; Chaiken & Eagly, 1983; Wood, Kallgren, & Preisler, 1985). Attitude depolarization is mediated by the process of associating heuristics with the validity of the message conclusion, for example, “experts said it so it must be true” (Todorov et al., 2002), or “length implies strength” (Chaiken & Stangor, 1987).

By contrast, recipients under high elaboration and thus following the central route to persuasion typically depolarize due to properties of thoughts generated in response to the

message source or persuasive message. That is, recipients generate thoughts in response to a persuasive message, and these thoughts may be evaluatively loaded – for example, in response to a persuasive appeal from a politician, a voter may think “this appeal does not seem genuine” (negative/unfavorable thought) or “this appeal would help many citizens” (positive/favorable thought). If recipients are low in thought favorability (i.e. generate a high number of unfavorable compared to favorable thoughts), they tend to depolarize (Greenwald, 1968; Love & Greenwald, 1978; Petty & Brock, 1981). This suggests that biasing the valence of thoughts by presenting unlikeable or untrustworthy message sources, for example, may lead to attitude depolarization.

In addition to thought favorability, attitude depolarization is mediated by meta-cognitive confidence (i.e. “thoughts about thoughts”), such as, the amount of confidence recipients place in the thoughts that they generate towards the attitude object (see Briñol & Petty, 2003; Petty, Briñol, & Tormala, 2002; Briñol et al., 2004; Tormala, Briñol & Petty, 2006, 2007). That is, some thoughts about attitude objects may be labelled as “more valid” than others, and this perceived validity subsequently influences one’s attitude – this is called the “self-validation hypothesis” (Petty et al., 2002). For example, during a political campaign a voter may think “this politician will change health policies for the better” (positive/favorable thought). However, the voter may place low confidence in this thought during later deliberation (e.g. “I am not confident that this politician will change health policies for the better”). Thought confidence is important because when confidence in thoughts is low, people tend to depolarize (Briñol et al., 2004; Petty, Briñol, & Tormala, 2002; Petty et al., 2007). This suggests that undermining confidence in people’s positive thoughts (or exacerbating confidence in people’s negative thoughts) following a persuasive intervention will result in attitude depolarization.

Recent work in confidence more generally (e.g. confidence in oneself) highlights the potential polarizing effects of high, as opposed to low meta-cognitive confidence. For instance, confidence in the extent to which one identifies with one's country, predicts extreme ingroup behavior, such as, willingness to "self-sacrifice" (Paredes, Santos, Briñol, Gomez, & Petty, 2019). Research examining the effects of non-verbal behavior on persuasion (see Guyer, Briñol, Petty, & Horcajo, 2019 for a review) corroborates this effect of confidence on attitudes and behavior. For example, nodding one's head during positive self-affirmation leads to greater physical performance in athletes (Horcajo, Paredes, Higuero, Briñol, & Petty, 2019). Moreover, recent work in self-validation shows that perceiving thoughts to originate from within oneself as opposed to an external source influences attitudes to a greater extent, due to greater perceived validity (Gascó, Briñol, Santos, Petty, & Horcajo, 2018). These findings suggest that self-validation processes may play a key role in self-persuasion as it does in the context of traditional persuasion, given that advocates generate persuasive messages themselves.

Summary

Traditional theories of persuasion, such as, dual-process models of persuasion, have significantly furthered our understanding of attitude and behavior change by making two key contributions: a) identifying classes of factors responsible for attitude depolarization (e.g. source, recipient, and persuasive message), and b) by identifying mechanisms (e.g. thought favorability, meta-cognitions) which may account for this depolarization. Collectively, the literature suggests that manipulating one or more of the tripartite factors (i.e. source, recipient, and message), will lead to attitude depolarization under certain circumstances. For instance, using a highly credible source to deliver a counter-attitudinal message containing high quality arguments is likely to facilitate depolarization among those who are high in Need-for-Cognition (i.e. those who enjoy thinking). Further, the literature suggests that

properties of thoughts drive attitude depolarization when people are actively processing message content. These include thought favorability, but also meta-cognitions, such as, thought confidence. This work suggests that thought favorability and thought confidence can be measured or manipulated and used to predict the extent of attitude depolarization during persuasive interventions.

However, traditional methods of persuasion are arguably limited in the context of bridging gaps between opposing groups. This is because attitude depolarization is difficult to achieve using standard persuasion when people are highly convicted in their attitudes (Krosnick & Petty, 1995). This means that while traditional persuasion continues to be fruitful in producing attitude change, it may be less effective in changing attitudes towards highly moralized and sacralized political issues, such as, migration and climate-change. These issues tend to produce greater partisan antipathy than those typically studied by traditional persuasion research (e.g. university policies, attitudes towards consumer products), and because partisans perceive each other's beliefs as "misguided" (Pew Research Center, 2014). Thus, novel methods of persuasion are required to induce attitude depolarization in a more subtle way.

Another limitation of traditional models is that they treat persuasion variables (message source, message, recipient) as mutually exclusive, and thus discount the potential that recipients have to induce attitude change themselves. That is, they assume that the message source and recipient are two different entities, when, in fact, as we demonstrate later, recipients themselves may act as message-generating sources. This may result in recipients persuading themselves (consciously or unconsciously) to adopt a new attitude position via self-persuasion (see Briñol et al., 2012; McCann, Higgins, & Fondacaro, 1991; Wilson, 1990). In this way, compared to traditional persuasion theories which focus on attitude change as a result of people processing a persuasive message, self-persuasion focuses

on attitude change as a result of people producing a persuasive message. Given that the persuasive message is generated by the recipient themselves, self-persuasion should result in less reactance compared to the recipient receiving an external persuasive message. Thus, we explore self-persuasion as a more viable method of depolarizing attitudes and encouraging more open communicative practices. In the following chapter, we explore emerging theories of persuasion, and demonstrate the value of self-persuasion as one possible method that can be used to mend our fractured attitudinal landscape.

Chapter 3 Emerging Theories of Persuasion

Introduction to Chapter 3

As highlighted previously, traditional methods of persuasion appear to be increasingly ineffective in producing attitude change, especially considering attitudes that are polarized and held with high confidence. An alternative to eliciting persuasion is via self-persuasion, whereby the message-generating source and recipient are the same person or group of people. Compared to traditional models of persuasion, whereby recipients passively receive messages from a message-generating source, self-persuasion models assume that recipients actively generate their own persuasive message. In this chapter, we explore the possibility of using self-generated messages (i.e. self-as-the-source) as a more viable means of depolarizing attitudes compared to standard persuasion, with the aim of increasing open communication between opposing groups.

In this non-comprehensive review, we examine the most commonly studied self-persuasion paradigms spanning psychology and communications research, including group polarization, mere thought, saying-is-believing, self-reflection, counter-attitudinal advocacy, and pro-attitudinal advocacy. By examining these paradigms more closely, we aim to provide a non-exhaustive list of factors which may drive attitude depolarization and open communicative practices between groups. In doing so, we provide recommendations for policymakers interested in designing interventions aimed at bridging gaps between polarized groups.

A Review of Self-Persuasion

Traditional research in persuasion examines how messages generated by a person or group (source) with persuasive intent, changes the attitudes of another person or group (recipient). Such persuasive attempts are often met with resistance and tend to be unsuccessful when persuasive intent is salient, or when recipient attitudes are extremely polarized (Cialdini, Petty, & Cacioppo, 1981; Petty & Cacioppo, 1979b). In the current socio-political climate, people tend to express highly polarized attitudes and tend to strongly advocate these attitudes. In order to effectively bridge gaps between polarized individuals, we consider an important alternative to standard persuasion, called self-persuasion (Briñol et al., 2012; McCann et al., 1991; Wilson, 1990).

According to theories of self-persuasion, attitude change can occur in the person generating the persuasive message, regardless of changes occurring in a hypothetical audience. Such alternatives to standard persuasion are especially important to examine in the current political climate, whereby people possess increasingly polarized views on important issues, such as, migration, and climate-change. Attitudes toward such issues tend to be highly moralized and therefore difficult to shift using standard persuasion. As such, we examine self-persuasion as an alternative and potentially more effective way of changing polarized attitudes. To-date, the process of self-persuasion has not been studied systematically – no comprehensive review currently exists in social psychology or communications research. The current chapter reviews the most common self-persuasion paradigms studied in these two literatures with the aim of uncovering specific factors which drive attitude depolarization.

Although the factors underlying depolarization in self-persuasion interventions are yet to be categorized systematically, the potential of these interventions to bridge gaps between social groups is much more evident. Consider a scenario in which a political conservative attempts to convince a liberal of his position on migration. As a result of this persuasion

attempt, the conservative may undergo unconscious self-persuasive processes, which may change the extremity of his initial attitude position or the confidence held this position. Importantly, the conservative may end up moving closer to, or further apart from the liberal on the attitude spectrum. That is, the self-persuasion effect can manifest in either polarization or depolarization of initial attitudes (Briñol et al., 2012; Janis & King, 1954; King & Janis, 1956; Sunstein, 2002; Tesser, 1976, 1978). Polarization occurs when attitude change takes place in the direction of the initial attitude, for example, a moderately anti-migration attitude becoming extremely anti-migration. On the other hand, depolarization occurs when attitude change occurs in the opposite direction to the initial attitude, for example, a moderately anti-migration attitude becoming neutral, or pro-migration. Uncovering factors which lead to attitude depolarization a) implies that these specific factors can be isolated and experimentally manipulated during offline and online self-persuasion interventions (e.g. discussion groups, chat platforms), and b) can be used to bridge political divides by stimulating civil conversation between attitudinally polarized groups.

Although a formal review of self-persuasion does not exist in the literature, a variety of literatures point to the underlying processes and factors which may be central to explicating attitude depolarization. In the following sections, we review each of these literatures briefly, point out key moderators of depolarization, and then provide a brief commentary on why more work is yet to be done in self-persuasion, more broadly.

Note, the aim of this review is not to meta-analyze or to systematically review all moderators of depolarization across these extensive and extremely diverse literatures. Rather, our aims are to: a) provide a non-exhaustive overview of the kinds of moderators that researchers and policymakers may be interested in when designing interventions aimed at moderating polarized attitudes, and b) highlight limitations of existing self-persuasion

paradigms as a basis for justifying the current research focus on pro-attitudinal advocacy (elaborated later in this chapter).

Group polarization. The classic investigation of self-persuasion in social psychology is group polarization (see Isenberg, 1986; Sunstein, 2002 for reviews). This work typically examines the impact of group discussion on individual choice preferences or acceptable levels of risk assigned to various activities (e.g. whether a chess-player should make a risky move, whether someone should cheat on their partner; see Burnstein & Vinokur, 1973, 1975, 1977; Lamm & Myers, 1978; Myers, 1975; Myers & Lamm, 1976; Stoner & Finch, 1961; Stoner, 1968; Zuber, Crott, & Werner, 1992). Advocates are first asked to indicate their preferences in private. Then, they are placed in groups and asked to arrive at a unanimous decision pertaining to a risk alternative or choice preference. Finally, advocates are asked to again indicate their personal choice preferences in private. In these studies, the average individual responses post-discussion tends to strongly resemble the group consensus ratings, as opposed to the pre-discussion ratings. This work demonstrates that exposure to group norms or interaction with groups can stimulate self-persuasion due to: a) social comparison (judging the social desirability of one's response by comparing the chosen alternative with that expressed by group members, or b) persuasive argumentation (considering persuasive arguments for each alternative injected by group members; see Burnstein & Vinokur, 1977; Myers & Lamm, 1976 for reviews).

According to social comparison theory, when tossing up between alternatives, people consider the social desirability of each alternative - during group discussion individual members compare his/her chosen alternative with those expressed by group members.

According to persuasive argumentation theory, when tossing up between alternatives, people consider the persuasive arguments in favor of each alternative. The group therefore exerts an informational influence by providing novel arguments that individual members may not have

considered previously. The self-persuasion effects occurring because of group discussion tend to be explained by both theories. Evidence does not unequivocally support one theory or the other, however, persuasive argumentation theory tends to be the favored explanation (Burnstein & Vinokur, 1977).

Nearly 60 years of work examining group polarization (first termed “risky shift”, see Stoner & Finch, 1961) shows that people tend to take riskier actions (i.e. display attitude polarization) following group discussion. However, a few streams of research demonstrate exceptions to this general finding. In seminal investigations of group polarization, the opposite effect to the typical “risky shift” (i.e. “cautious shift”, akin to depolarization) was observed when advocates discussed issues such as, marriage, on which society praises cautiousness (Stoner, 1968). This effect appears to be moderated by knowledge of others’ choices. Specifically, knowing what others choose in a two-choice task leads advocates to question why they selected the alternative instead (Burnstein & Vinokur, 1975). This knowledge leads to depolarization, as indexed by a greater preference for the alternative choice post-discussion.

Depolarization also occurs if novel persuasive arguments supporting the opposing opinion are presented to the group, as it forces advocates to consider arguments that they have not considered previously (cf. persuasive argumentation; Burnstein & Vinokur, 1977). Attitudes which are highly moralized (e.g. attitudes towards capital punishment) tend to depolarize as result of people considering novel persuasive arguments, although, these highly moralized attitudes depolarize to a lesser extent compared to lukewarm attitude issues, such as, choice of product alternatives (Vinokur & Burnstein, 1978).

Much more recently, depolarization has been shown to occur in both online and offline platforms in the presence of a moderator (Strandberg, Himmelroos, & Grönlund, 2017). That is, depolarization occurs even in like-minded groups when a facilitator is present

to reinforce conversational norms, express appreciation for reasoned justifications, and to emphasize principles of respect and sincerity. Conversely, corroborating seminal work, polarization tends to occur when unmoderated discussion is allowed between like-minded group members, regardless of initial stance towards the issue (Strandberg et al., 2017). This finding suggests that the presence of an active moderator may stimulate depolarization even in attitudinally homogenous groups.

Although the group polarization effect tends to resemble self-persuasion at first glance, the inferences that can be drawn from this work are limited. First, group polarization research is typically concerned with choice preferences, not attitudes. Second, although the main contention is that persuasive arguments injected into the group drive attitude change, it is unclear as to what aspects of group discussion actually facilitate (de)polarization. The method of measuring group polarization (post-discussion ratings minus pre-discussion ratings) does not allow us to pinpoint people's shifts in attitudes. That is, there is no way of knowing the processes involved in shifting a person's position on a particular dimension. In a group context, we cannot disentangle standard persuasion processes and self-persuasive processes. Attitude change could be driven by social comparison, persuasive arguments via standard persuasion, or self-persuasion via a different unidentified process. Finally, conclusions about groups may not necessarily apply to individuals. Even if overall change in group-level attitudes occurs, each group member themselves may have not changed substantially (Fraser, Gouge, & Billig, 1971).

Mere thought. Contrasting group polarization, simply thinking about an attitude object can lead to polarization, even in the absence of a group (mere thought effect; Leone & Ensley, 1986; Sadler & Tesser, 1973; Tesser, 1976; Tesser & Conlee, 1975; Tesser & Cowan, 1977; Tesser, Leone, & Clary, 1978). In mere thought paradigms, participants are asked for their attitude towards a target person or object, and then given an opportunity to silently

reflect on the attitude object. The handful of studies conducted on mere thought tend to suggest polarization, overall, as a result of thinking about the attitude object (see Tesser, 1978 for a review). This is likely because people tend to generate attitude-consistent thoughts during deliberation, such that if people like the object they are more likely to generate favorable thoughts about the object (Clarkson, Tormala, & Leone, 2011; Clarkson et al., 2013; Tesser & Leone, 1977). The longer the time spent deliberating about the attitude object, the more the attitude polarizes, regardless of initial extremity or direction (Tesser & Conlee, 1975; Tesser & Cowan, 1977).

Nevertheless, a closer examination of the literature suggests a few exceptions to this general finding. For instance, when a distraction is introduced, less attitude polarization occurs (Sadler & Tesser, 1973). Conversely, when the attitude object is made salient (i.e. the advocate is asked to think about the attitude object again as opposed to thinking about something unrelated), greater polarization results. This finding suggests that introducing a distraction during the thinking process may attenuate polarization³.

Some evidence suggests that thinking with “reality constraints” (i.e. the object present in sight) similarly attenuates polarization. For example, thinking without reality constraints (i.e. object not present in sight) has been shown to produce greater attitude polarization (Tesser, 1976)⁴. By extension, the authors suggest that thinking with reality constraints will attenuate polarization⁵. Other research shows that allowing less time for advocates to think, increases the probability of depolarization (Tesser & Conlee, 1975). For instance, reducing time spent thinking from 3 minutes to 30 seconds results in a 27% increase in the probability of depolarization (Study 1; Tesser & Conlee, 1975).

³ This effect was greater for those who were initially more positive towards the attitude object (a target person) suggesting that holding initial negative attitudes may override the effects of distraction (see Sadler & Tesser, 1973).

⁴ This occurred for females only, with paintings as the attitude object.

⁵ This idea has been supported by studies conducted in counter-attitudinal advocacy, which is discussed later in this paper (see Tuller et al., 2015).

More recent studies conducted on mere thought have demonstrated similar effects of time spent thinking. Less time spent thinking (60 seconds vs 180 seconds) results in greater depolarization compared to moderate time spent thinking (Clarkson, Tormala, & Leone, 2011), suggesting that reducing the time allowed for advocates to think may induce depolarization. Interestingly, the relationship between depolarization and time spent thinking resembles a U-shaped curve, such that high time spent thinking (300 seconds) leads to a spike in levels of depolarization (Clarkson et al., 2013). This is likely due to exhaustion of attitude-consistent thoughts, which may then reduce confidence in thoughts generated (cf. thought confidence; Clarkson et al., 2013). Thus, for those who generate more attitude-inconsistent thoughts compared to attitude-consistent thoughts, increasing thought confidence via a false feedback mechanism results in attitude depolarization as well (Clarkson et al., 2013). This suggests that manipulating thought confidence by undermining the quality of thoughts advocates generate may drive attitude depolarization.

Saying-is-believing. The communications literature suggests that speakers often modify their messages to suit their listeners – this effect is known as audience tuning or audience design (Bell, 1984; Clark & Carlson, 1982; Clark & Murphy, 1982). In designing messages for their audiences, speakers tend to take into account their audience's knowledge and attitudes towards the communication topic (Clark & Haviland, 1977; Fussell & Krauss, 1989a, 1989b; Higgins, 1992; Higgins & Rholes, 1978; McCann, et al., 1991). In a typical saying-is-believing paradigm, participants read and summarize a description of an attitude object (usually a target person) for a target who either ostensibly likes or dislikes the attitude object. Then, participants indicate their own attitudes towards the attitude object and are asked to re-create its original description. Interestingly, describing attitude object information to another person allows reinterpretation of ambiguous information to match the recipient's attitude. For instance, speakers who are led to believe that their audience is favorable towards

an attitude object, design their message description to contain more positively valenced words (Fussell & Krauss, 1989a, 1989b).

Consequently, this audience design results in a saying-is-believing effect (SIB) whereby the speaker's own attitude towards the target is distorted in the direction of the audience (see Higgins, 1992 for a review). Closer examination of the literature reveals an important qualification to the SIB effect - audience attitudes do not directly impact speaker attitudes, unless speakers actually use message descriptions that are evaluatively consistent with audience attitudes (Higgins & McCann, 1984). That is, the effect of audience attitudes on speaker attitudes is mediated by the valence of the message description. To illustrate, when the audience attitude towards the target is positive, and if the speaker uses more positive descriptions of the target (compared to negative or neutral descriptions) the speaker's attitude towards the target will become more positive as well (akin to attitude polarization)⁶. This suggests that asking speakers to write balanced descriptions (both positive and potential negative descriptions of the target) may attenuate attitude polarization.

In addition to the valence of the message description, there are other factors which moderate the SIB effect, such as, the size of the target audience, and whether speakers receive feedback from their audience (Echterhoff, Higgins, & Groll, 2005; Echterhoff, Higgins, Kopietz, & Groll, 2008; Echterhoff, Higgins, & Levine, 2009; Hausmann, Levine, & Higgins, 2008). To illustrate, the SIB effect does not occur for live three-person audiences compared to single-person audiences (Hausmann et al., 2008). This holds true unless the speaker receives feedback that the three-person audience is able to accurately identify the described target. This implies that for larger audiences, validation is required for attitude polarization to occur as a result of audience design. This suggests that polarization via SIB

⁶ This effect is stronger for high authoritarians compared to low authoritarians, and even stronger when the audience possess high status, suggesting some moderation by personality variables (see Higgins & McCann, 1984).

can be attenuated by: a) increasing the size of the live audience, and by b) withholding audience validation.

Corroborating work on audience validation, more recent research shows that the SIB effect is moderated by speakers' perceptions of common ground with their audience - this holds true in contexts in which speakers formulate judgements of a particular target, and then generalize this judgement to an ambiguous new target which shares some similarities to the initial target (Bebermeier, Echterhoff, Bohnet, & Hellmann, 2015). Interestingly, the SIB effect generalizes to an ambiguous new target only when speakers are told that they shared some common judgements with their audience. This implies that attitude polarization and generalizations occurring via SIB can be attenuated if: a) speakers are led to believe that they do not possess commonalities with their audience, or if b) the new ambiguous target is sufficiently distinguished from the initial target.

Although research in audience design demonstrates the importance of target in self-persuasion, the key difference between our work and audience design, is that the latter examines communication to enable comprehensibility, as opposed to persuasion. However, audience design is important because it highlights the importance of the audience in driving persuasion: different kinds of messages tailored to different audiences can bring about different extents of self-persuasion. Audience design highlights that generating messages can lead to self-persuasion, and that the extent of self-persuasion can occur to varying degrees depending on who you are talking to. The saying-is-believing literature notes multiple audience characteristics which may impact speaker attitudes, such as, whether audience feedback is received or not, the size of the audience, and most commonly, the valence of audience attitudes (Echterhoff et al., 2005, 2008, 2009; Hausmann et al., 2008). However, it is important to note that in the context of self-persuasion, variation in audience attitudes has not always produced variation in self-persuasion (Briñol et al., 2012). We examine this

finding in more detail later in this chapter, but flag that audience attitudes may not be as central to self-persuasion as some other factors, such as the content of the message generated.

Self-reflection. According to work on self-reflection, generating reasons to justify an attitudinal position brings about attitude change, typically in the form of depolarization (Nisbett & Wilson, 1977; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984; Wilson & Schooler, 1991; Wilson & Dunn, 1986; Wilson, Kraft, & Dunn, 1989; Wilson, 1990). In these studies, participants are asked to generate reasons for their attitude towards a target person or object. Depolarization is measured by evaluating the direction and extremity of attitudes pre versus post self-reflection.

Similar to work on the mere thought effect, attitudes depolarize following a self-reflection intervention depending on the valence of reasons generated (see Wilson, 1990 for a review). Generating more positive reasons results in more favorable object evaluations, however, certain attitudes (such as, attitudes toward art) tend to be grounded in reasons that are difficult to verbalize overtly. Advocates may experience difficulty explicating reasons for these attitudes, and may even generate evaluatively inconsistent reasons, resulting in the adoption of a counter-attitudinal position by endorsing the new attitude implicated by their biased reasons (Wilson & Schooler, 1991). As such, similar to the mere thought paradigms, undermining thought confidence via a false feedback mechanism (cf. Clarkson et al., 2011, 2013) may undermine confidence in reasons generated, and thus the drive attitude depolarization.

However, notably, work in self-reflection shows that analyzing reasons for one's attitude results in attitude depolarization only if people possess little knowledge about the attitude object. For instance, likeability ratings of political candidates were influenced by a significant reasons and knowledge interaction, such that significant attitude change consistent with the valence of reasons only occurred for unknowledgeable participants (Wilson et al.,

1989). Whether this effect is because knowledgeable people have stronger attitudes, or because knowledgeable people have a better understanding of their reasons for their attitudes is unknown, and therefore, may confound the results. As such, although we know that generating reasons to justify attitudes tends to drive attitude change, and that low knowledgeability impacts the extent of this change, based on the work in self-reflection, we cannot infer how these effects are driven.

Counter-attitudinal advocacy. Unlike mere thought, self-reflection and saying-is-believing paradigms, the effects of communication with persuasive intent (as opposed to comprehensibility) has been demonstrated by counter-attitudinal advocacy paradigms, in which participants are asked to generate arguments for a position opposite to their own (see Hovland et al., 1953 for a review). In this paradigm, advocates are asked to either deliver a talk to an audience, or to write arguments supporting the opposing viewpoint (“active” condition); read counter-attitudinal arguments (“passive” condition); or to write arguments supporting an unrelated position (“control” condition). The general finding in seminal research is that depolarization occurs only when advocates actively generate counter-attitudinal messages as opposed to passively reading counter-attitudinal arguments (Hovland, et al., 1953; Hovland, Lumsdaine, & Sheffield, 1949; Hovland & Mandell, 1952; Kelman & Hovland, 1953; Sherif, Hovland, & Harvey, 1957).

Multiple explanations for this effect have been proposed, including biased scanning theory, (attitudes shifting due to “active” participants searching for one-sided arguments; Janis & Gilmore, 1965); cognitive dissonance theory (attitudes shifting due to motivation to restore internal consistency; Aronson, 1969; Festinger, 1962; Festinger & Carlsmith, 1959), and self-perception theory (inferring attitudes from overt behavior; Bem, 1967). Further, similar to mere thought and self-reflection effects, thought favorability may influence depolarization. That is, generating more negative (attitude-inconsistent) thoughts than

positive (attitude-consistent) thoughts during advocacy may lead to attitude depolarization (Greenwald, 1968; Greenwald & Albert, 1968; Love & Greenwald, 1978; McGuire, 1968). In addition, properties of the arguments generated may impact depolarization, for example, perceiving one's own arguments to be more original than those generated by others may account for counter-attitudinal advocacy effects (Greenwald, 1969). All these explanations carry equal merits and validity in explaining depolarization (Petty & Cacioppo, 1981), although dissonance theory tends to take precedence in most research exploring counter-attitudinal advocacy (see Kim, Allen, Preiss, & Peterson, 2014 for a recent review).

Interestingly, counter-attitudinal advocacy effects tend to be moderated by issue familiarity or knowledgeability, such that advocates depolarize on unfamiliar attitude issues (e.g. the duration of a common cold), but not on familiar attitude issues (e.g. those related to entertainment; Janis & King, 1954). Further, depolarization appears to be moderated by the extent to which advocates improvise their counter-attitudinal arguments (King & Janis, 1956). Specifically, rates of depolarization are higher when advocates improvise counter-attitudinal arguments after reading a pre-written script (guided improvisation) compared to simply reading aloud the script. Moreover, guided improvisation tends to produce greater depolarization compared to unguided improvisation, during which participants independently generate counter-attitudinal arguments from scratch (Matefy, 1972). This suggests that the extent of improvisation moderates depolarization, such that complete counter-attitudinal improvisation may produce back-fire effects.

In addition to improvisation, a recent meta-analysis of 230 studies was conducted to examine the effects of a range of dissonance-arousing moderators including perception of choice (whether advocates felt that they were engaging freely or forcibly in advocacy), task commitment (whether advocates needed to put extra effort into advocacy, such as, through improvisation or delivering the message to an audience), and private versus public advocacy

(whether advocates delivered the message to an audience or not; see Kim, Allen, Preiss, & Peterson, 2014). The authors predicted that the greater the dissonance aroused (during improvisation as opposed to non-improvisation, during free choice advocacy, during greater task commitment and public delivery) the greater the attitude depolarization via self-persuasion.

Interestingly, no single moderator exerted a significant effect on its own, but the overall effect of counter-attitudinal advocacy on depolarization was significant ($r = .23$). This finding points to the historical difficulty of explaining counter-attitudinal advocacy effects, and suggests that all dissonance-arousing interventions may be similarly effective in producing attitude depolarization (Kim et al., 2014).

Much more recent work shows that depolarization is moderated by previously unconsidered factors, namely, whether advocates: a) meet a disagreeing opponent prior to advocating in private, and b) are held accountable for their message (Tuller, Bryan, Heyman, & Christenfeld, 2015). Depolarization occurs when advocates briefly affiliate with someone they later discover holds an opposing opinion on the target issue, but not when advocates generate arguments for an abstract person holding the opposing view. Additionally, this effect of affiliation is qualified by whether advocates are held accountable for their message – that is, whether they believe that their message will be read by the person that they met previously. This is likely because advocates expend greater effort in writing counter-attitudinal arguments that will be viewed by someone who is qualified to scrutinize their work⁷. In sum, even brief affiliation with someone holding an opposing view (e.g. having a casual conversation on topics unrelated to politics) may induce depolarization if people are held accountable for their advocacy attempt.

⁷ Depolarization does not occur when advocates are told that their arguments will be read by “an expert”, suggesting that the accountability effect is further moderated by the attitude of the intended audience.

Positive effects of affiliation have been supplemented by very recent work investigating counter-attitudinal advocacy and value-congruence. It appears that taking the perspective of the opposition may back-fire and lead to attitude polarization, unless the opposition shares similar values to the advocate (e.g. those related to political ideology; Catapano, Tormala, & Rucker, 2019). Similar to work on common ground and the saying-is-believing effect, this suggests that highlighting similarities in values between opponents may attenuate possible back-fire effects of perspective-taking interventions.

Despite its value, the most obvious limitation of this work is the ecological validity of counter-attitudinal advocacy, in that people do not typically spontaneously advocate for an opposing point of view. In this way, counter-attitudinal advocacy paradigms are characterized by forced compliance. Our aim is to find an indirect way of changing attitudes, and to encourage people to question their attitudes without directly asking them to do so or forcing them to consider an opponent's position. This limitation in ecological validity is circumvented by the most recent self-persuasion paradigm: pro-attitudinal advocacy, which explores self-persuasion within the context of people advocating for their own point of view.

Pro-attitudinal advocacy. Unfortunately, most work in advocacy focuses on counter-attitudinal advocacy, despite greater external validity and ubiquity of pro-attitudinal advocacy in contemporary society (see Briñol et al., 2012; Gordijn et al., 2001; Tuller et al., 2015 for a few exceptions). In contrast to counter-attitudinal advocacy, advocates in a pro-attitudinal advocacy intervention generate arguments for their own position on an attitude issue. Attitude depolarization is measured either within-subjects (post-advocacy attitudes minus pre-advocacy attitudes) or between-subjects (post-advocacy attitudes between two conditions). Note, pro-attitudinal advocacy interventions are importantly different to group polarization interventions, as advocacy is undertaken in private as opposed to in public. Thus, self-

persuasion is a more likely explanation for depolarization during pro-attitudinal advocacy, as opposed to effects of both self-persuasion and external persuasion in group settings.

One particular moderator of depolarization identified in pro-attitudinal advocacy contexts is numerical support for one's position. Specifically, advocates depolarize more frequently following an advocacy episode when they are told that they hold the minority opinion as opposed to the majority opinion (Gordijn et al., 2001). This finding suggests that emphasizing that the advocate possesses the minority opinion (if true) prior to advocacy may lead to attitude depolarization due to concerns about social validation ("consensus implies correctness" heuristic; Axson et al., 1987). This effect of numerical support on attitude depolarization is further qualified by ingroup-outgroup dynamics, such that advocates exhibit greater depolarization via perceptions of threat when they are led to believe that their own position is a political outgroup's majority position. Taken together, this work suggests that: a) numerical support is a persuasion variable which may be used to influence attitude depolarization, and b) perceptions of threat is mechanism by which attitude depolarization may occur.

Another moderator identified in the context of pro-attitudinal advocacy is target of persuasion. Compared to counter-attitudinal advocacy, pro-attitudinal advocacy leads to depolarization when the target of persuasion is oneself as opposed to another person (Briñol et al., 2012). This is mediated by meta-cognitive confidence advocates place in their advocacy attempt, such that advocates perceive that: a) they put less effort into convincing themselves as opposed to another person, and b) they perceive the arguments aimed at themselves as weaker than those aimed at another person. Taken together, these findings suggest that target characteristics is a persuasion variable which may be used to influence attitude depolarization and that confidence in advocacy (as indexed by subjective perceptions of effort and argument quality) is a mechanism by which depolarization may occur.

Note how confidence in advocacy is similar to confidence in thoughts, which is captured by meta-cognitions. Similarly, confidence in one's advocacy attempt is captured by meta-cognitive confidence (see Petty et al., 2007 for a review) by measuring subjective perceptions of one's advocacy attempt. Importantly, meta-cognitive confidence includes not just properties of thoughts (as per the traditional definition of meta-cognitions), but other content relevant cues, such as, perceptions of the amount of effort expended in generating arguments (perceived effort; cf. Briñol et al., 2012), confidence in the strength of one's arguments (perceived argument quality; cf. Briñol et al., 2012), and confidence in one's credibility or expertise to generate the message (perceived knowledgeability or source credibility; see Ehret, Van Boven, & Sherman, 2018; Rios et al., 2018). Importantly, low meta-cognitive confidence in one's advocacy attempt as indexed by one or more of the above variables may lead to attitude depolarization and subsequently more open communicative practices, such as, seeking attitude-inconsistent information or expressing receptivity to opposing views.

Interestingly, although target effects exist in pro-attitudinal advocacy, self-persuasion effects do not vary when target is varied by type of audience (i.e. colleague, friend, etc., see Briñol et al., 2012). This suggests that other factors, such as content of the message generated may take precedence over target in accounting for attitude depolarization and open communicative practices following pro-attitudinal advocacy. We will explore this possibility in detail in the chapters to follow.

In sum, although pro-attitudinal advocacy is extremely understudied, it provides a fruitful avenue to subtly influence polarized attitudes. People advocate for their opinions on a regular basis, and there are numerous channels readily available for advocacy (e.g. Reddit).

Thus, the current research seeks to explore: a) specific manipulations of pro-attitudinal advocacy may which predispose advocates to self-persuasion effects, such as,

encouraging advocates to generate practical as opposed to moral messages, b) the downstream consequences of pro-attitudinal advocacy on attitude depolarization and communicative practices (e.g. confirmation bias, receptivity to opposing views, intentions to proselytize), and c) the processes (e.g. self-persuasion, motivations, language used) underlying pro-attitudinal advocacy which may lead to such practices.

The Current Research

Existing research in pro-attitudinal advocacy has made a few key contributions to our understanding of the factors and mechanisms responsible for attitude depolarization. Despite the decades of research undertaken in counter-attitudinal advocacy (CAA), this thesis focuses on the self-as-the source in the context of pro-attitudinal advocacy (PAA), during which advocates generate persuasive arguments for their own point of view (see Briñol et al., 2012; Gordijn et al., 2001). This is because PAA is: a) more ecologically valid compared to CAA, as people are more likely to advocate for their own attitude as opposed to someone else's attitude, and b) highly relevant to the current socio-political climate, as online platforms provide ample opportunity for people to proselytize their own opinions. In spite of its potential to mend the currently fractured attitudinal landscape, PAA has received little scholarly attention – a gap in the literature this thesis aims to address.

The following limitations emerge from the scant existing literature: a) attitude depolarization has been measured by comparing between-subjects post-advocacy attitudes, as opposed to within-subjects pre-post attitude change, which limits inferences which can be drawn at an individual-level, b) attitude issues studied tend to be less controversial (e.g. attitudes towards university policies) and unrepresentative of issues currently debated in the political climate, c) only meta-cognitive confidence has been systematically explored as a potential mechanism for self-persuasion effects occurring during PAA, d) only attitude depolarization has been explored as the outcome of pro-attitudinal advocacy, although

behavioral intention measures should correspond more strongly with the adoption of open communicative practices, and e) only numerical support and target characteristics have been explored as factors or persuasion variables that can be manipulated to produce attitude depolarization in experimental settings.

Given these limitations, the current research expands upon the pro-attitudinal advocacy literature in the following important ways:

First, we strictly use a pre-post design to measure attitude (de)polarization within-subjects to infer clearly whether self-persuasion has occurred for each participant, as opposed to simply on a group-level. We also test whether depolarization and changes in subsequent communicative intentions occurs for controversial socio-political and moral issues pertinent to the current political climate, such as, migration and climate-change.

Second, we move beyond meta-cognitive confidence-induced self-persuasion as the only mechanism by which PAA exerts its effects on attitude depolarization and communicative practices. We show that although meta-cognitive confidence is a potent mechanism for attitude depolarization (Abeywickrama & Laham, 2020a; see Chapter 4), it is not the only mechanism by which PAA can bridge attitudinal gaps. For instance, we show that moral expressiveness (advocates' tendency to express words related to moral values) predicts attitude depolarization as well (Abeywickrama, Rhee, Crone, & Laham, 2020; see Chapter 5).

Third, we expand beyond attitudes and attitude depolarization as the primary dependent variable, to capturing consequences of advocacy on between-person interactions. We conceptualize this suite of behavioral intentions as open versus closed communicative intentions. In doing so, we focus on a variety of phenomena that capture openness verses closedness in interactions, such as, seeking information that is consistent or inconsistent with one's attitudes (confirmation bias/congeniality bias; see Hart et al., 2009; Nickerson, 1998 for

reviews), tendency to persuade others of one's opinion (proselytization; see Cheatham & Tormala, 2015, 2017), intentions to engage with the opposition, and choosing interaction partners with similar or dissimilar views (see Rios et al., 2018). To illustrate, we show that manipulating style of advocacy (i.e. Attack vs Defense) predicts information-seeking preferences and willingness to choose interaction partners holding dissimilar views (Abeywickrama & Laham, 2020b; see Chapter 6).

Extant research provides further clues as to how communicative practices such as, information-seeking will be influenced in the context of pro-attitudinal advocacy. For example, research in selective exposure shows that defense motivation (motivation to confirm one's existing views) leads to greater congeniality bias compared to accuracy motivation (motivation to learn new information; see Hart et al., 2009 for a review). This suggests that persuasion variables which influence accuracy motivation may similarly predict positive information-seeking tendencies ("open" communicative practice). On the other hand, extant work in attitude certainty shows that high confidence or certainty in one's attitudes leads to greater proselytization intentions ("closed" communicative practice; see Cheatham & Tormala, 2015, 2017). Therefore, in the context of pro-attitudinal advocacy, we predict that similar to attitude confidence, high meta-cognitive confidence in one's advocacy attempt should lead to greater proselytization intentions. This expansion of dependent variables beyond attitude depolarization is important given that attitudes need not change in order to influence behavior (see attitude-behavior correspondence; Ajzen & Fishbein, 2005). We derive support for this tenet in our empirical studies (see Chapters 5 & 6).

Finally, we extend existing literature by testing whether certain persuasion variables can change one's proclivity towards attitude depolarization, congeniality bias, and proselytization, for example. One such persuasion variable is message content (generating moral versus practical arguments; see Chapter 5), and a second is style of advocacy

(attacking opponent's position (Attack) versus defending one's own position (Defense); see Chapter 6).

We show that changes to message content (moral versus practical arguments) influences attitude depolarization and intentions to proselytize one's attitudes, not via traditional meta-cognitive processes, but via moral expressiveness (see Chapter 5). Specifically, advocates encouraged to write only practical arguments for their position are more likely to depolarize and less likely to proselytize their opinions, compared to those encouraged to write only moral arguments supporting their position. This effect is mediated by the moral content of the message, such that those generating only practical arguments use a significantly lower number of words related to core moral values, such as, harm, fairness, and purity (see moral foundations theory; Graham et al., 2013 for a review). As such, we demonstrate that manipulations of persuasion variables, such as message content, may influence open communicative practices by changing the way in which advocates express their attitudes.

Further, we show that changes to style of advocacy (degrading the opposition versus justifying one's own position) directly influences information-seeking tendencies, without influencing attitude depolarization (see Chapter 6). More specifically, degrading the opposing position (countering opponent's arguments) leads to uncongeniality bias compared to justifying one's own position. This is possibly due to heightened accuracy motivation induced by the realization that advocates do not possess full knowledge of the opponent's arguments. In this way, we provide support for the notion that attitude depolarization is not a necessary pre-requisite for advocates to demonstrate more open communicative practices following pro-attitudinal advocacy.

Concluding Remarks

In this thesis, we exclusively focus on PAA, given its ubiquity in the current socio-political climate, and its untapped potential to bridge gaps between polarized groups. In doing so, we acknowledge that self-persuasion is a primary mechanism for attitude depolarization, but it is not the only mechanism by which open communicative practices can be encouraged. We consider the possibilities of moral expressiveness and motivations as direct or indirect mediators of the effects of PAA on communicative intentions. Further, in addition to target effects, we explore the effects of different types of persuasion variables, such as, message content and style of advocacy on attitudinal and behavioral outcomes. Thus, our primary research question is as follows:

What factors and mechanisms in the context of pro-attitudinal advocacy lead to attitude depolarization and open communicative practices?

We began in Chapter 3 with a review of the factors that lead to attitude depolarization via self-persuasion. In Chapter 4, we undertake an empirical test of meta-cognitive confidence-induced self-persuasion in the context of pro-attitudinal advocacy, whereby participants advocate their attitudinal position freely. However, as suggested, it is possible that different persuasion variables (such as message content and style of advocacy) affect depolarization and communicative practices such as, congeniality bias, and proselytization intentions. Thus, in Chapters 5 and 6 we test the effects of message content (moral versus practical arguments) and advocacy style (attacking opponent's position versus justifying one's own position). Finally, in Chapter 7 we conclude with a General Discussion of the theoretical and practical implications of our findings, and end with a commentary on harnessing the power of pro-attitudinal advocacy in bridging ideological divides.

Chapter 4 Meta-cognitive Confidence, Need-for-Cognition, and Self-Persuasion

Preface to Chapter 4

(This chapter was previously published in *Social Psychology* as “Meta-Cognition Predicts Attitude Depolarization and Intentions to Engage with the Opposition Following Pro-Attitudinal Advocacy”; the accepted version of the article is included in this thesis)

The key goals of this research stream were to: a) determine whether pro-attitudinal advocacy (PAA) may lead to attitude depolarization, even for attitudes towards controversial socio-political issues, and b) explore the processes and boundary conditions of attitude depolarization during PAA. Specifically, we explored meta-cognitive confidence (confidence in one’s thoughts and attitude-relevant judgements; see Petty & Briñol, 2014 for a review), as a key mechanism for attitude depolarization in the context of PAA, with the hypothesis that attitudes will depolarize when meta-cognitive confidence is low.

This is because previous research in both standard and self-persuasion suggests low meta-cognitive confidence as a driver of attitude depolarization. In the context of standard persuasion, low confidence in one’s thoughts (i.e. low thought confidence) has been shown to predict attitude depolarization (Petty et al., 2002; Petty et al., 2007). For example, following the presentation of a message about a university policy, participants reported less favorable attitudes towards the policy if they possessed low compared to high thought confidence (Petty et al., 2002). Similarly, in the context of self-persuasion, perceiving one’s arguments to be of a low quality has been shown to predict attitude depolarization (Briñol et al., 2012). For example, following advocacy for a university policy, participants report less favorable attitudes towards the policy if they perceive their arguments to be weak compared to strong. Drawing on this work, we hypothesized that meta-cognitive confidence will similarly predict

attitude depolarization in the context of pro-attitudinal advocacy. In addition, we extended the exploration of meta-cognitive confidence effects beyond attitude measures only, to include communicative intentions (e.g. willingness to engage with dissimilar others), given that behavioral intentions may change even in the absence of attitude change (Ajzen & Fishbein, 2005).

In addition to exploring confidence-polarization and confidence-intention links, we explored level of elaboration as a boundary condition to these effects. We examined Need-for-Cognition (NfC) as a proxy for level of elaboration (see Cacioppo et al., 1996 for a review). Previous work has demonstrated links between NfC and meta-cognitive confidence, such that those high in NfC are more likely to use meta-cognitive confidence to inform their attitudes (Tormala et al., 2002), and revise their attitudes if they experience low confidence (Wegener & Petty, 1997). Thus, we hypothesized that NfC will moderate confidence-polarization and also confidence-intention links, such that those who are high in NfC will be more likely to depolarize and engage with dissimilar others when meta-cognitive confidence is low.

Citation: **Abeywickrama, R. S., & Laham, S. M. (2020a).** Meta-Cognition Predicts Attitude Depolarization and Intentions to Engage with the Opposition Following Pro-Attitudinal Advocacy. *Social Psychology*. <https://doi.org/https://doi.org/10.1027/1864-9335/a000424>

**Meta-Cognition Predicts Attitude
Depolarization and Intentions to Engage with
the Opposition Following Pro-Attitudinal
Advocacy**

Introduction

Attitude polarization has undoubtedly increased over the last decade (Pew Research Center, 2014). A pressing question is therefore posed: is it possible to depolarize attitudes with the aim of bridging gaps between opposing groups? Traditional attempts at depolarization whereby a source sends a persuasive message to recipient, tend to be unsuccessful, especially when advocates place high confidence in their attitudes (Krosnick & Petty, 1995). Thus, we explore the possibility of *self-persuasion* (cf. Briñol, McCaslin, & Petty, 2012) as an alternative route to attitude depolarization.

Previous work in self-persuasion largely supports a “saying-is-believing” effect, whereby people’s initial attitudes become more extreme (i.e. polarize) following episodes of communal elaboration, such as, discussing one’s attitude in a group (group polarization; see Isenberg, 1986; Sunstein, 2002 for reviews), or describing an attitude object to someone else (audience design; Higgins & Rholes, 1978). Interestingly, simply thinking about an attitude object in private (mere thought effect; see Tesser, 1978 for a review), or reflecting on reasons for holding an attitude (self-reflection; see Wilson, 1990 for a review) has been shown to result in attitude polarization as well.

However, some research suggests that people’s attitudes become less extreme (i.e. depolarize) following *counter-attitudinal advocacy*, that is, advocating for a position discrepant to one’s own (Briñol et al., 2012; Greenwald & Albert, 1968; Hovland, Janis, & Kelley, 1953; Janis & King, 1954; King & Janis, 1956; Tesser, 1978; Tuller, Bryan, Heyman, & Christenfeld, 2015; Zimbardo, 1965). One reason why counter-attitudinal advocacy brings about depolarization is because when advocates generate arguments in support of an opposing position, they become less confident in their own stance due to cognitive dissonance (see Kim, Allen, Preiss, & Peterson, 2014 for a review of dissonance-arousing processes during counter-attitudinal advocacy).

The majority of research on advocacy and self-persuasion centers around counter-attitudinal advocacy (see Briñol et al., 2012; Gordijn, Postmes, & de Vries, 2001; Mercier & Landemore, 2012; Tuller et al., 2015 for exceptions). Yet, it remains unclear whether attitude depolarization occurs when people advocate for their *own opinions* (i.e. pro-attitudinal advocacy; Briñol et al., 2012). Given that people are generally more likely to advocate for their own attitudes as opposed to others, we seek to explore self-persuasion effects as a result of pro-attitudinal advocacy. Further, given the potential role of confidence in advocacy effects, we explore the possibility that confidence will drive self-persuasion effects during pro-attitudinal advocacy.

Meta-Cognition and Self-Persuasion

Two fundamental questions guide the current research: a) Can pro-attitudinal advocacy sometimes lead to attitude depolarization? And if so, b) What drives depolarization, and under what conditions does it occur? Given previous work in standard persuasion⁸ literature, we might expect the effects of pro-attitudinal advocacy on attitude depolarization to be moderated by meta-cognitive properties of the advocacy attempt. Research in standard persuasion shows attitude change to be driven by *meta-cognitive confidence*, i.e., how confident people are in their thoughts, attitudes, judgements or beliefs following exposure to a persuasive message (see Petty & Briñol, 2014; Petty et al., 2007; Wagner, Briñol, & Petty, 2012 for reviews). Notably, low meta-cognitive confidence induced by a persuasive message predicts depolarization. For example, when people generate positive thoughts in response to a persuasive message but are not confident in those thoughts, they tend to depolarize (Petty, Briñol, & Tormala, 2002). Importantly, meta-cognitive confidence does not only apply to

⁸ Here “standard persuasion” refers to traditional methods of persuasion, whereby a source delivers a persuasive message aimed at a particular recipient (see Heuristic-Systematic Model and Elaboration Likelihood Model of persuasion for more details; Chaiken & Trope, 1999; Chen & Chaiken, 1999; Petty & Briñol, 2014; Petty & Briñol, 2012). Standard models of persuasion thus imply that the source and recipient are two separate entities. On the other hand, in models of self-persuasion, the message-generating source and recipient are one entity; i.e. the message is generated and received by the *same* person or group of people (see Briñol et al., 2012; Isenberg, 1986; Tesser, 1978; Wilson, 1990 for reviews).

thoughts, but to other attitude-relevant constructs as well (Petty et al., 2007). For instance, if people are not confident in the arguments that they generate, they tend to depolarize (Briñol et al., 2012).

Research in self-persuasion suggests a plethora of meta-cognitive drivers of attitude depolarization. These include favorability of thoughts generated in response to advocacy attempts (Love & Greenwald, 1978), confidence in thoughts (Clark, Wegener, Sawicki, Petty, & Briñol, 2013; Clarkson, Tormala, & Leone, 2011), aspects of the quality of arguments generated, such as, originality and strength (Greenwald & Albert, 1968), and perceptions of effort expended in generating arguments (Briñol et al., 2012). To illustrate, impressions of one's weakness in advocacy (indexed by a combination of perceived effort, and perceived quality of arguments) predicts attitude depolarization following pro-attitudinal advocacy (Briñol et al., 2012). This work collectively suggests that meta-cognitive confidence is a key driver of self-persuasion effects during advocacy. Study 1 will test this possibility.

In addition to attitude depolarization, we expand the effects of pro-attitudinal advocacy and meta-cognitive confidence to other relevant dependent variables, such as, communicative intentions. This is because although attitudes held with high conviction may not depolarize as a result of advocacy (Krosnick & Petty, 1995), advocates may nevertheless display more "open" communicative intentions if they lack confidence in their advocacy attempts. Thus, as a result of advocating one's position, one may become more receptive to the opinions of dissimilar others (Voelkel & Feinberg, 2018). On the other hand, even if advocacy does not change one's attitudes, it may make one more willing to persuade others of one's position (Cheatham & Tormala, 2015, 2017). In Study 2 we test these contrasting possibilities.

Effects of Need-for-Cognition

We now turn our attention to the boundary conditions under which the relationships between meta-cognitive confidence, depolarization and communicative intentions might be exacerbated or attenuated. A potential moderator of our effects that may give insight into the mechanism by which meta-cognitive confidence predicts advocacy-related dependent variables, is *level of elaboration* (Petty & Cacioppo, 1984, 1986). Standard persuasion research identifies level of elaboration (i.e. amount of thinking, effort expended) as a key factor impacting attitude change. Increased deliberation during persuasive message processing tends to lead to greater attitude change due to favorable message-relevant thoughts, especially when messages consist of high-quality arguments presented by credible sources (see Briñol & Petty, 2015; Petty & Cacioppo, 1986 for reviews). One predictor of level of elaboration, is Need-for-Cognition (NfC), which is an individual pre-disposition towards enjoying and engaging in effortful cognitive activity (Cacioppo et al., 1984, 1986).

NfC has received particular interest in the context of standard persuasion for many reasons. Compared to low NfC individuals, attitudes of high NfC individuals tend to be formed by central cues (e.g. argument quality) as opposed to peripheral cues (e.g. source attractiveness); high NfC individuals tend to generate more meta-cognitions, such as, attitude-relevant thoughts compared to low NfC individuals, and finally, attitudes formed by high NfC individuals are more durable and predictive of behavior compared to those formed by low NfC individuals (see Cacioppo, Petty, Feinstein, & Jarvis, 1996 for a comprehensive review of NfC).

However, the effects of NfC in the context of self-persuasion, are less well-known. We explore whether NfC correlates with self-persuasion effects, such as, attitude polarization, or whether meta-cognitive confidence – self-persuasion links are instead informed by heuristic processes. Specifically, we consider moderation by NfC, and explore

whether greater or less elaboration predicts the relationships between meta-cognitive confidence, attitude depolarization, and communicative intentions. We hypothesize that the greater tendency one has to think about one's advocacy attempt (high NfC), the more likely one is to turn lack of confidence in advocacy (low meta-cognitive confidence) into depolarization and intentions to engage with those holding opposing views. That is, we suggest that NfC may attenuate the effect of meta-cognitive confidence on attitude polarization when meta-cognitive confidence is low.

We put forward two key arguments in support of this hypothesis. First, regardless of whether high NfC individuals experience increased meta-cognitive confidence (see Petty, Briñol, Loersch, & McCaslin, 2009 for a review), they are more likely to use confidence to inform their subsequent attitudes. Those high in NfC have a higher propensity to think about the validity of their thoughts (Briñol & Petty, 2003), and thus are more likely to base their judgements on meta-cognitive cues compared to those low in NfC (Tormala, Petty, & Briñol, 2002). This implies that high NfC individuals will be more open to revising their beliefs and thus display greater attitude depolarization when meta-cognitive confidence is low.

Second, NfC has recently been shown to moderate the link between subjective experiences of difficulty generating arguments (i.e. low meta-cognitive confidence) and attitudes. A recent study asked participants to generate either many or few counter-arguments against a particular attitude issue (Walter & Cohen, 2019). Those who were high in NfC displayed more favorable attitudes (i.e. were persuaded more) after experiencing difficulty counter-arguing, because they relied on their meta-cognitive cues to a greater extent in formulating their judgements, compared to those who were low in NfC. Those who were low in NfC displayed less favorable attitudes (i.e. were persuaded less) when they experienced low meta-cognitive confidence in their advocacy attempt. By extension then, those who are high in NfC should display less favorable attitudes (i.e. be persuaded more, or depolarize)

after experiencing difficulty arguing for their *own* attitude (pro-attitudinal advocacy), again, as they rely more on their meta-cognitive experiences, compared to low NfC individuals. In addition to displaying attitude depolarization, high NfC individuals should then also be more open to interacting with the opposition when meta-cognitive confidence in advocacy is low, given that they are more likely to proactively start discussions with others (Henningsen & Henningsen, 2004) and revise their beliefs when confidence is low (Wegener & Petty, 1997).

The Current Research

In the current research, we conduct preliminary correlational studies to examine self-persuasion effects by simulating real-world online advocacy contexts in which people direct pro-attitudinal advocacy attempts at an audience who disagrees with them. In Study 1 we measure a suite of previously tested meta-cognitive variables to explore their impact on attitude depolarization. In Study 2, we attempt to replicate the confidence – polarization link with a larger sample, but with two key modifications: a) expansion of dependent variables to include not only attitude polarization, but also communicative intentions, such as, intentions to engage with disagreeing others, and intentions to persuade others, and b) inclusion of NfC as a potential moderator of confidence-self-persuasion links⁹.

⁹ In Study 2 we also examined variations in pro-attitudinal advocacy by manipulating target audience (i.e. participants were allocated to one of three advocacy conditions with neutral, disagreeing or agreeing hypothetical target audiences). Given that we did not find a significant effect of target, we omit details in the main body of text. See footnote 5 below for additional detail on this manipulation, and a more general treatment of this aspect of the design in the Supplementary Materials.

Study 1

Method

Participants and design. First-year psychology students ($N = 321$, $M_{\text{age}} = 19.67$, $SD = 4.17$, 17 – 55 years, 240 female) participated in exchange for course credit. Our sample size was sufficient to detect small effects in correlational analyses ($|r| = .20$) with 95% power, (GPower programme; Faul, Erdfelder, Buchner, & Lang, 2009), which is the effect size typically found in social psychology (Richard, Bond, & Stokes-Zoota, 2003).

This research was cleared by the Human Research Ethics Committee (HREC) at the University of Melbourne and adheres to the highest ethical standards set by the Declaration of Helsinki. All participants provided informed consent prior to participation and were clearly informed of their right to withdraw from the study, and that withdrawal will not carry any penalty, grades or otherwise. Participants were informed that their confidentiality would be protected, and that de-identified data would be made publicly available via the Open Science repository.

Materials and procedure. Participants completed an online survey administered via Qualtrics (Qualtrics, 2016). Participants first indicated their attitudes towards a range of issues (e.g. migration, consumer products, carbon emissions policy) and also the confidence they placed in these attitudes. The attitude issue of interest was embedded within these attitude measures to avoid suspicion. The target issue was a hypothetical carbon emissions policy which the government may institute in Australia.

Time 1 attitude extremity. Participants were first presented with the text below:

“Carbon dioxide emissions from electricity and gas usage contribute to more than half of Victoria’s total carbon dioxide emissions. These emissions peak particularly during colder times of the year, due to significantly increased usage of heating, for example. The Australian

government may pass a law specifying that households set a maximum thermostat temperature of 20 degrees Celsius during winter”.

Participants reported the extent to which they agreed with the carbons emissions policy on a 7-point Likert scale ($1 = \text{strongly disagree}$; $7 = \text{strongly agree}$; $M = 4.27$, $SD = 1.83$).

Time 1 attitude confidence. Participants then reported the confidence they placed in this attitude on two 9-point scales (“How certain are you of your attitude towards the policy?” and “How confident are you of your attitude towards the policy?” ($1 = \text{not at all}$, $9 = \text{very much}$; taken from Tormala & Petty, 2002). Responses were averaged to form a composite score ($r = .78$; $M = 6.39$, $SD = 1.63$).

Participants were next asked to generate persuasive arguments to convince someone who disagreed with them on their attitude towards this policy.

Thought-listing. Following advocacy, participants listed the thoughts they had in response to engaging in this task (maximum eight thoughts; taken from Clark, Wegener, Sawicki, Petty, & Briñol, 2013).

Time 2 attitude extremity and confidence. For a second time, participants indicated their attitude towards the policy ($M = 4.22$, $SD = 1.88$) and how confident they felt in their attitude ($r = .92$; $M = 6.71$, $SD = 1.67$). Participants then completed the following thought-related meta-cognitive measures in a randomized fashion: thought favorability, thought confidence, and thought conviction.

Thought favorability. To assess thought favorability, participants were presented with the thoughts that they generated previously and asked to rate each thought with one of the following options: *positive*, *negative*, *neutral* or *unrelated to the issue*. A thought favorability index was computed by subtracting the number of negative thoughts from the number of

positive thoughts and dividing this difference score by the total number of issue-relevant thoughts generated (see Clark et al., 2013; $M = -.06$, $SD = .76$).

Thought confidence. Participants were asked to rate the confidence they placed in the thoughts listed previously using four 9-point scales (e.g. “Overall, how much confidence do you have in the thoughts that you listed?”, “Overall, how valid would you say your thoughts are?”; taken from Clark et al. 2013). These items were averaged to create a reliable composite of thought confidence (Cronbach’s $\alpha = .87$; $M = 5.89$, $SD = 1.76$).

Thought conviction. Participants indicated the conviction they placed in their thoughts using two 9-point scales (“My thoughts in response to the issue of a maximum thermostat temperature are connected to my core beliefs/ reflect something stable/enduring about who I am” (taken from Abeywickrama & Laham, 2017). Responses to these items were averaged to create a reliable composite measure of thought conviction ($r = .79$; $M = 5.62$, $SD = 2.01$).

Next, participants rated the perceived quality of arguments generated, and the amount of effort expended in generating their arguments (random order).

Perceived argument quality. Participants rated the perceived strength of their arguments on three 9-point scales (“How strong are these arguments?”, “How valid are these arguments?”, “How satisfied are you with these arguments?”; taken from Briñol et al., 2012). These items were averaged to form a composite score of perceived argument quality (Cronbach’s $\alpha = .91$; $M = 5.72$, $SD = 1.83$).

Perceived level of effort. Participants indicated the amount of effort they put into generating their arguments on three 9-point scales, “How much attention did you pay to the task of generating arguments?”, “How deeply did you think about the issue when generating arguments?”, “How much energy did you put into generating your arguments?”; taken from

Briñol et al., 2012). Items were averaged to create a reliable composite score (Cronbach's $\alpha = .89$; $M = 5.72$, $SD = 1.68$).

Finally, participants responded to a demographics questionnaire (age, gender).

Results and Discussion

Arguments generated by participants were checked for being on topic and adherence to the manipulation and no exclusions were made on this basis.

Time 1 attitude distribution. Attitude scores were transformed into a bipolar scale (-3 to +3) for ease of interpretation and further computation. A bimodal distribution was obtained for attitudes at Time 1. Based on the attitude distribution, participants were clustered into three “initial stance” groups: “Pro” ($n = 155$; those who responded with transformed scores of +1, +2, or +3), “Anti” ($n = 121$; those who responded with -1, -2, or -3), and “Neutral” ($n = 45$, those who responded with 0 on the 7-point scale). Only Pro and Anti initial stances were included in the main analyses, given that participants in the Neutral condition either argued for the policy or against the policy (not necessarily *pro*-attitudinal advocacy), leaving a final sample size of $N = 276$.

Attitude depolarization. A polarization index was calculated to account for: a) direction (depolarize, polarize or unchanged), and b) extent of attitude change. We created a score for each participant to reflect whether they depolarized (-1), polarized (+1) or remained unchanged (0). This score was multiplied by the extent of change (i.e. change of 1, 2, 3, 4, 5, or 6 scale-points from Time 1 to Time 2) to form an overall polarization index for each participant.

This specific index to study depolarization was created based on seminal research undertaken to examine (de)polarization in the context of self-persuasion (see Tesser & Leone, 1977). This index provides a more accurate indication of attitude change post-advocacy, as it accounts for both extremity and direction of attitude change, compared to pre-post difference scores, for example. For instance, a participant with T1 attitude (-3) and T2 attitude (-1),

would have a polarization score of (-2), because they depolarized by two scale points, while a participant with T1 attitude (+1) and T2 attitude (+3) would receive a polarization score of (+2), because they polarized by two scale points instead.

A single sample t-test revealed that all participants significantly depolarized, on average, following the advocacy task ($M = -.33$, $SD = 1.27$, $t(275) = -4.25$, $p < .001$, Cohen's $d = .26$). This depolarization was not significantly predicted by whether people were initially for or against the policy ($\beta = .00$, $t(274) = -.14$, $p = .89$)¹⁰.

Predictors of attitude depolarization. Given that multiple measures of meta-cognition were positively correlated with each other (see Table 4-1 in the Supplementary Materials), we conducted a principal components analysis to examine the separability of all our meta-cognitive variables (perceived argument quality, thought confidence, thought favorability, thought favorability*thought confidence composite¹¹, thought conviction and perceived effort). A principal components analysis with a direct oblimin rotation extracted a two-component solution. The pattern matrix with rotated component loadings placed thought confidence (.84), argument quality (.80), thought conviction (.76) and effort (.51) onto a single component. Thought favorability (-.99) and the index of thought favorability and thought confidence (-.98) strongly loaded onto a separate component. Accordingly, the factor solution represents two dimensions of meta-cognition, the first resembling general *meta-cognitive confidence*, and the second, *thought valence*.

Component scores (Regression method) on each of these components were computed for each participant. Both components (meta-cognitive confidence and thought valence) were added as simultaneous predictors into a regression model containing initial stance, absolute initial attitude extremity (absolute attitude extremity, regardless of direction), and Time 1

¹⁰ Note: β = standardized regression coefficients, significance level $\alpha = .05$.

¹¹ Given that confidence in positive versus negative thoughts may have different effects on attitude depolarization, we multiplied thought favourability by thought confidence to create a general measure of confidence in thoughts. Positive scores on this measure indicate greater confidence in positive thoughts.

attitude confidence¹². The overall regression model significantly predicted depolarization ($R^2_{\text{Adjusted}} = .13$, $F(5, 253) = 8.36$, $p < .001$), while meta-cognitive confidence marginally significantly predicted depolarization, such that, lower meta-cognitive confidence predicted greater depolarization ($\beta = .12$, $p = .055$). On the other hand, thought valence failed to uniquely predict depolarization ($\beta = -.08$, $p = .19$), suggesting overall meta-cognitive confidence as the key driver of depolarization in the advocacy context (see Table 4-2 in the Supplementary Materials).

Given our imperfect measure of thought favorability*thought confidence, we re-ran a principal components analysis excluding this measure. This analysis revealed a single meta-cognitive confidence component explaining 46% of the variance, which significantly predicted depolarization ($\beta = .14$, $p = .02$) in a subsequent regression model containing initial stance, initial absolute attitude extremity and Time 1 attitude confidence ($R^2_{\text{Adjusted}} = .13$, $F(4, 254) = 10.38$, $p < .001$; see Table 4-3 in the Supplementary Materials for the full regression table).

Concluding remarks

In sum, Study 1 sought to explore: a) if pro-attitudinal advocacy may lead to attitude depolarization in a pre-post design, and b) the role of meta-cognition in accounting for depolarization effects. On an issue related to climate-change, we show that somewhat counter-intuitively, attitudes may depolarize via self-persuasion following pro-attitudinal advocacy. This depolarization is predicted by meta-cognitive confidence in one's advocacy attempt, such that the lower the meta-cognitive confidence, the greater the extent of

¹² Absolute attitude extremity significantly predicted attitude depolarization, such that, the more extreme the attitude, the greater the depolarization ($\beta = -.37$, $p < .001$). To circumvent this problem, previous work in self-persuasion tends to only focus on moderately positive or negative attitudes (e.g. Tesser, 1978). However, given that we are interested in depolarizing extreme attitudes, participants on the two extremes were retained in our analyses. In addition, we find that Time 1 attitude confidence significantly predicted depolarization, such that, attitudes held with low confidence at Time 1 predicted greater depolarization following advocacy ($\beta = .22$, $p = .001$). This is consistent with work on attitude confidence (for a review see Tormala, 2016), however, tangential in terms of the current research as we are particularly interested in meta-cognitive moderators of self-persuasion effects. The missing cases are due to not all participants reporting issue-relevant thoughts during advocacy (which affects the thought favorability index).

depolarization. Importantly, attitude depolarization was not predicted by initial attitudinal position, suggesting that the results are generalizable to advocates on either side of the attitudinal spectrum.

Study 2

In Study 2, we: a) attempt to replicate the correlation between meta-cognitive confidence and depolarization, b) extend the scope of dependent variables beyond attitudes to behavioral intentions to engage with and persuade others, and c) explore Need-for-Cognition (NfC) as a potential moderator of the relationships between meta-cognitive confidence, attitude depolarization and communicative intentions, to examine if these relationships are related to advocates' level of elaboration¹³. Specifically, we predict that the relationships between: a) meta-cognitive confidence and depolarization, and b) meta-cognitive confidence and communicative intentions, will be moderated by NfC, such that, high NfC will predict depolarization and greater intentions to engage with the opposition when meta-cognitive confidence is low.

Method

Participants and design. Five-hundred and fifty-nine undergraduates (17-55 years, $M = 19.30$, $SD = 2.99$, female = 412) participated in exchange for course credit. This sample size yield 99% power to detect a similar effect size as Study 1 ($|r| = .20$). However, in the context of multiple regression, given that power depends on collinearity between predictors, we acknowledge that there is an upper limit to how much power we have in detecting

¹³ It has been suggested that *target of advocacy* also likely matters for self-persuasion, as people may engage differently in convincing others who vary in their initial attitude position (Briñol et al., 2012). In support of this view, the communications literature provides evidence of *audience design*, whereby speakers take into account audience characteristics in designing their utterances (Bell, 1984; Clark & Murphy, 1982) and that this impacts speakers' own attitudes toward the topic (Echterhoff, Higgins, & Groll, 2005; Echterhoff, Higgins, Kopietz, & Groll, 2008; Echterhoff, Higgins, & Levine, 2009; Higgins & Rholes, 1978). Thus, in Study 2 we also considered whether our effects are influenced by audience attitudes (more direct test of "saying-is-believing") by randomly allocating participants to direct their advocacy attempts at a hypothetical audience who either agrees with them, disagrees with them, or is neutral towards the attitude issue. Similar to Briñol and colleagues (2012) we did not find this to be the case (see Supplementary Materials for full details).

significant effects. Nevertheless, the sample size is sufficient to detect even smaller effects ($|r| = .12$) with 80% power¹⁴.

Materials and procedure. The study followed a similar procedure to Study 1, other than for a few exceptions detailed below:

Time 1 attitude extremity and confidence. Similar to Study 1, participants indicated their attitude towards the carbon emissions policy ($M = 4.25$, $SD = 1.79$). Those who were neutral towards the issue were re-directed a separate study ($n = 64$), leaving $N = 495$ as the final sample size. Participants then indicated the confidence they placed in their attitude ($M = 5.38$, $SD = 1.68$).

Advocacy task. As a filler task, and to reduce suspicion, participants were asked to indicate their attitudes towards a variety of issues (e.g. banning plastic bags at grocery stores, lowering the minimum driving age, mandatory military service). They then advocated for their position on the carbon emissions policy¹⁵.

Following the main task, participants completed the thought-listing task, and then indicated their Time 2 attitude ($M = 4.36$, $SD = 1.77$) and attitude confidence ($M = 6.37$, $SD = 1.77$). Participants then completed thought favorability ($M = .10$, $SD = .70$), general thought confidence ($M = 5.91$, $SD = 1.43$) and thought conviction ($M = 5.94$, $SD = 1.76$) measures in a random order.

Next, participants completed additional measures of meta-cognitive confidence, including perceived argument quality ($M = 5.39$, $SD = 1.54$), perceived effort expended in

¹⁴ The initial power analysis was conducted to enable detection of small-medium effects ($f = .15$) in a 2(Initial stance: Pro vs Anti) x 3(Target: Agree, Disagree, Neutral) between-subjects design with 80% power. However, given that this analysis yielded no significant result of target, we have reported the results of this analysis in the Supplementary Materials.

¹⁵ Note that at this stage participants were randomly allocated to generate arguments aimed at one of three hypothetical audiences: someone who agrees with them (Agree, $n = 167$), disagrees with them (Disagree, $n = 165$), or someone neutral towards the issue (Neutral; $n = 163$; see Supplementary Materials for full details). Similar to Briñol and colleagues (2012), given we did not find variation depending on target attitudes, we analyze all cases together.

generating arguments ($M = 5.69$, $SD = 1.42$). An additional measure assessing perceptions of the self as a knowledgeable source (perceived source credibility) was included. This measure included two 9-point scales (*not at all to very much*) on the following items; “To what extent do you think you are a credible source to discuss this kind of issue?” and “To what extent do you think you are a knowledgeable to discuss this kind of issue?”; adapted from Clark et al., 2013). These items were averaged to create a composite score (Cronbach’s $\alpha = .84$; $M = 3.99$, $SD = 1.74$).

Finally, participants responded to behavioral intention measures capturing “open” and “closed” communicative practices in a randomized order.

Intentions to engage with opposition. Open communicative practice was assessed by asking participants their intentions to engage with the opposition using three 9-point scales (“To what extent are you willing to listen to/talk to/negotiate with someone holding an opposite viewpoint on this topic?”; 1 = *not at all*, 9 = *very much*; Cronbach’s $\alpha = .81$; $M = 6.32$, $SD = 1.68$).

Persuasion intentions. Closed communicative practice was assessed by asking participants the likelihood of persuading others of their position using three 9-point scales (“How likely would you be to try to persuade your family or friends/someone you do not know well but see often/a stranger, of your own position on this topic?”; 1 = *not at all*, 9 = *very much*; adapted from Cheatham & Tormala, 2015). Items were averaged to create a composite score of persuasion intentions (Cronbach’s $\alpha = .86$; $M = 4.89$, $SD = 1.83$).

Need-for-Cognition (NfC). Finally, participants completed a shortened version of NfC (see Epstein, Pacini, Denes-Raj, & Heier, 1996). Participants rated the extent to which each of five statements (e.g. “I prefer complex to simple problems”) applied to themselves on a 5-point Likert scale (1 = *completely false*, 5 = *completely true*). Standard and recoded

reverse items were averaged to create a composite score (Cronbach's $\alpha = .70$, $M = 17.46$, $SD = 3.54$).

Results and Discussion

Attitude depolarization. Similar to Study 1, roughly equal numbers of advocates were for and against the policy ($n_{\text{Anti}} = 226$, $n_{\text{Pro}} = 269$). Consistent with the results of Study 1, a single sample t-test revealed that all advocates significantly depolarized, on average, following the advocacy task ($M = -.31$, $SD = 1.10$, $t(494) = -6.32$, $p < .001$). Interestingly, an effect of initial stance emerged, such that, those who were initially against the carbon emissions policy depolarized to a greater extent than those who were initially favorable ($M_{\text{Anti}} = -.46$, $SD = 1.25$, $M_{\text{Pro}} = -.19$, $SD = .94$, $p < .001$; see Supplementary Materials for full details). Thus, we control for initial stance in the analyses that follow.

Meta-cognitive confidence and attitude depolarization. As in Study 1, a principal components analysis was performed on all the meta-cognitive variables (perceived argument quality, thought confidence, perceived source credibility, perceived effort, thought conviction, thought favorability; see Table 4-4 in the Supplementary Materials for the correlation matrix). The analysis extracted a single component solution explaining 48% of the variance. The component matrix revealed argument quality (.84), perceived source credibility (.75), thought confidence (.77), perceived effort (.71), thought conviction (.67) as the highest loading variables. Thought favorability loaded less strongly onto the overall confidence component (.26).

Component scores (Regression method) were computed for the overall meta-confidence component, and added into a regression model containing initial stance, initial attitude extremity, and initial attitude confidence (see Table 4-5 in the Supplementary Materials). The overall regression model significantly predicted depolarization ($R^2_{\text{Adjusted}} = .09$, $F(4, 472) = 13.37$, $p < .001$). Replicating Study 1, the meta-cognitive confidence

component significantly predicted depolarization, such that, lower meta-cognitive confidence predicted greater depolarization following advocacy ($\beta = .15, p = .002$).

Meta-cognitive confidence and communicative intentions. Interestingly, not only does meta-cognitive confidence predict attitude depolarization, but it also influences engagement and proselytization. Intentions to engage and intentions to persuade were moderately highly correlated ($r = .34$), thus were statistically controlled for in the regression analyses below. Specifically, meta-cognitive confidence strongly predicted intentions to proselytize one's attitudes ($\beta = .46, p < .001$) in a regression model containing initial stance and intentions to engage with the opposition, such that lower the meta-cognitive confidence in one's advocacy attempt, the less likely one is to proselytize. Similarly, controlling for intentions to proselytize, lower meta-cognitive confidence predicted lower intentions to engage with the opposition ($\beta = .29, p < .001$). This is somewhat counter-intuitive, because we might expect to find lower meta-cognitive confidence predicting greater intentions to engage with the opposition. This is because engaging with the opposition presents advocates with the opportunity to seek novel information, to clarify doubts, and to form more informed attitudes. However, as noted previously, the relationship between meta-cognitive confidence and subsequent attitude polarization and communicative intentions is likely moderated by level of elaboration.

Need-for-Cognition, meta-cognitive confidence and communicative intentions.

Given their chronic tendency to enjoy effortful cognitive activity and preference for more complex compared to simple message processing, individuals who are high in NfC should generally display increased meta-cognitive confidence in their advocacy attempt. Standardized values (z-scores) were computed for the NfC measure. Consistent with the above prediction, NfC positively predicted meta-cognitive confidence, controlling for initial stance in one's position ($t(474) = 2.65, \beta = .12, p < .01$).

Next, we explore NfC as a moderator of the relationships between: a) meta-cognitive confidence and depolarization, and b) meta-cognitive confidence and communicative intentions.

We conducted a moderation analysis on PROCESS (Hayes, 2013), using 5000 bootstrapped samples with 95% confidence intervals and the same covariates used in Study 1 analyses (initial stance, initial attitude extremity and initial attitude confidence; see Table 4-6 in the Supplementary Materials for full regression results). First, in exploring the links between meta-cognitive confidence and (de)polarization, we see that NfC moderates the effect of meta-cognitive confidence on (de)polarization in a manner that suggests that confidence-polarization effects are at least partially dependent upon elaboration ($\Delta R^2 = .01$, $F(1, 470) = 4.67$, $p = .03$). When NfC is one standard deviation below the mean, there is no relationship between increasing meta-cognitive confidence on attitude polarization ($B = .07$, $SE = .07$, 95% CI $[-.07, .20]$, $p = .33$)¹⁶. However, when NfC is at the mean ($B = .17$, $SE = .05$, 95% CI $[-.07, .27]$, $p < .001$) or one standard deviation above the mean ($B = .27$, $SE = .07$, 95% CI $[-.13, .41]$, $p < .001$), there is a significant relationship between meta-cognitive confidence and attitude polarization.

Importantly, when meta-cognitive confidence is low (one standard deviation below the mean), those who are high in NfC (one standard deviation above the mean) depolarize more compared to those who are lower in NfC (see Figure 4-1). For those who are low in NfC, low meta-cognitive confidence translates into increased polarization compared to those who are high in NfC. Although the results are trending in this particular direction, they are not statistically significant.

¹⁶ B = unstandardized regression coefficients as generated in PROCESS, significance level $\alpha = .05$.

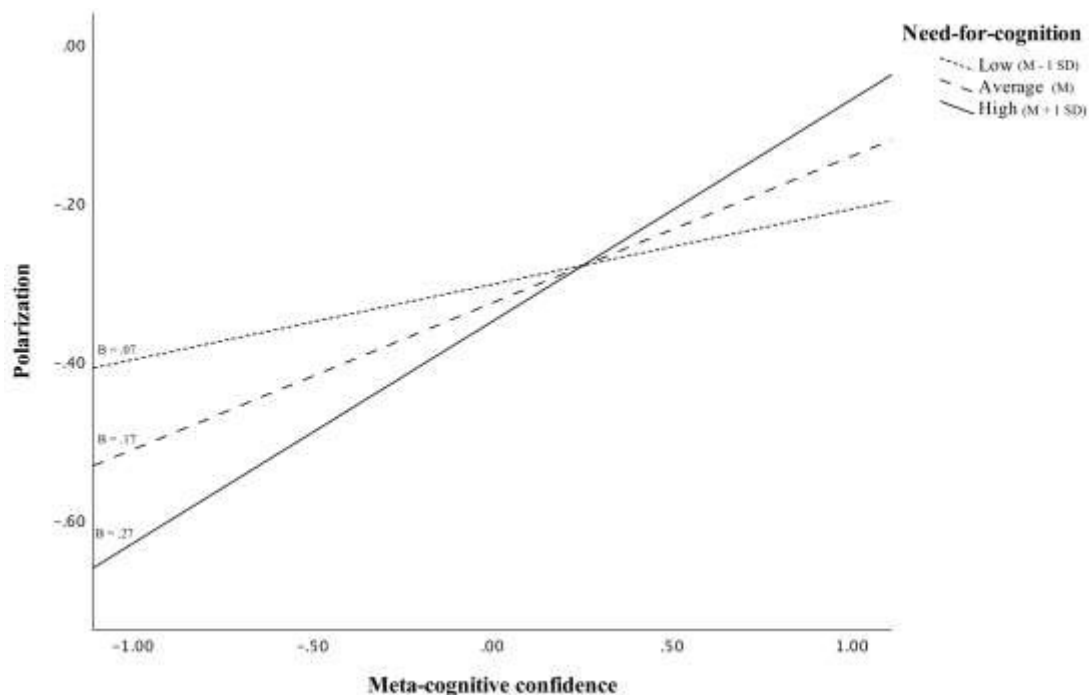


Figure 4-1. Relationship between meta-cognitive confidence and polarization at different levels of Need-for-Cognition

Next, we examined the possibility that NfC moderates the relationship between meta-cognitive confidence and communicative intentions (see Table 4-7 in the Supplementary Materials for full regression results). A moderation analysis examining the link between meta-cognitive confidence and communicative intentions shows that NfC significantly moderates the link between meta-cognitive confidence and intentions to engage with the opposition, controlling for intentions to persuade others ($\Delta R^2 = .02$, $F(1, 469) = 14.12$, $p < .001$). Specifically, when NfC is low, there is a strong relationship between meta-cognitive confidence on engagement intentions ($B = .40$, $SE = .10$, 95% CI[.19, .60], $p < .001$), and no relationship when NfC is moderate ($B = .15$, $SE = .08$, 95% CI[-.02, .31], $p = .08$) or high (B

= -.10, $SE = .11$, 95% CI[-.32, .11], $p = .36$). Notably, when meta-cognitive confidence in advocacy is low (one standard deviation below the mean) those who are high in NfC are more willing to engage with those holding opposing views ($B = .74$, $SE = .10$, 95% CI[.54, .95], $p < .001$), compared to those who are low or average in NfC (see Figure 4-2). However, for those who are low in NfC, low meta-cognitive confidence translates into less willingness to interact with the opposition ($B = .24$, $SE = .09$, 95% CI[.06, .42], $p = .01$).

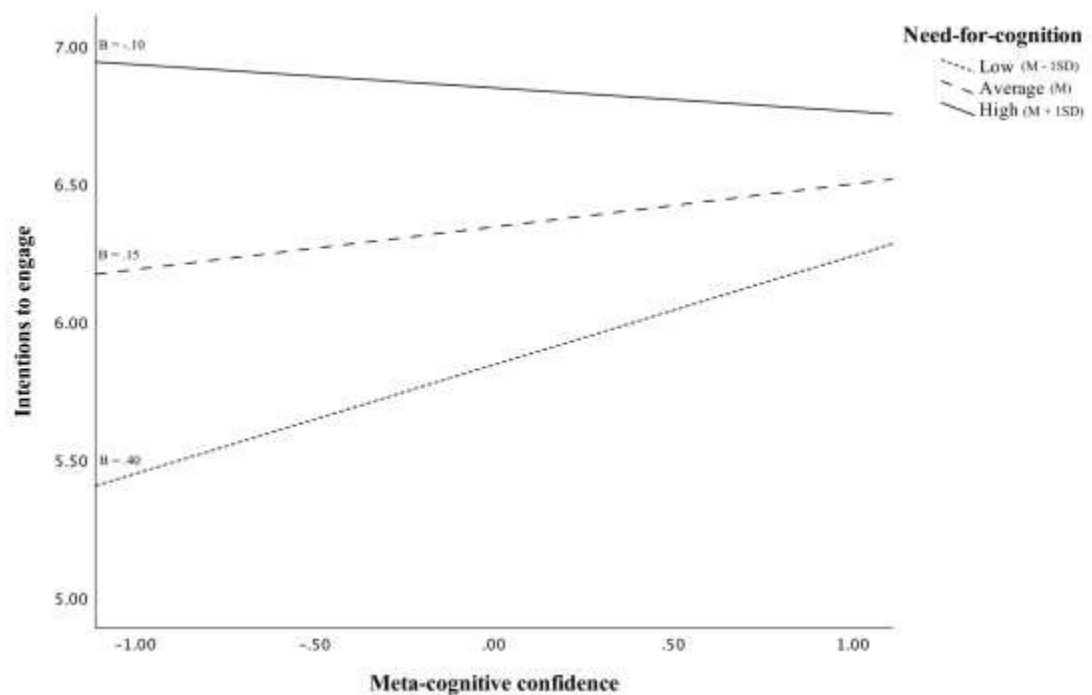


Figure 4-2. Relationship between meta-cognitive confidence and intentions to engage with the opposition at different levels of Need-for-Cognition

By contrast, controlling for intentions to engage with the opposition, NfC does not moderate the strong relationship between meta-cognitive confidence and intentions to persuade others ($\Delta R^2 = .00$, $F(1, 469) = .10$, $p = .75$).

Taken together, the moderation analyses suggest that NfC may contribute to open practices by translating lack of confidence in advocacy into open communicative intentions. For those who are not inclined to elaborate (low NfC), lack of confidence may translate into less depolarization and less engagement with the opposition (i.e. decreased open practices). The results suggest that while NfC leads to increased confidence in one's advocacy attempt, on average, it may also lead to "opening up" of communicative practices by increasing willingness to interact with the opposition, when confidence is low. The results suggest that high NfC may be a "protective" factor against closedmindedness in advocacy contexts, by stimulating approach tendencies towards dissimilar others.

General Discussion

Across two studies we show that people may depolarize during pro-attitudinal advocacy, and that depolarization is driven by the meta-cognitive confidence people place in their advocacy attempts. We also show that in the current work, depolarization is unlikely to depend on external factors, such as, target attitudes (as shown by the lack of audience effects outlined in the Supplementary Materials). In addition to predicting depolarization, we show that meta-cognitive confidence also predicts communicative intentions, such as intentions to proselytize and engage with the opposition. If people lack confidence in their advocacy attempts, they are less likely to proselytize, and more likely to engage with the opposition. The latter opening effect of meta-cognitive confidence on communicative practices partly depends upon one's tendency to elaborate, as indexed by NfC. Lack of confidence translates into greater depolarization and increased intentions to engage with dissimilar others for those who are more likely to elaborate on their thoughts (high NfC). These results suggest that the

opening effect of meta-cognitive confidence on communicative practices may be dependent upon elaboration, suggesting that meta-cognitive confidence – self-persuasion links may be driven by deliberative as opposed to heuristic processes.

Meta-Cognitive Confidence, Depolarization and Communicative Intentions

It may be argued that the results of the current study are influenced by repeated attitude expression due to the pre-post design employed. However, previous work studying the effects of pre-post designs have yielded mixed results concerning the effects of repeated attitude expression on attitude extremity. While some work shows that repeated expression leads to more extreme attitudes (i.e. attitude polarization; see Downing, Judd & Brauer, 1992), others have shown no relationship between repeated attitude expression and attitude extremity, but rather, a positive relationship between repeated expression and attitude accessibility (Powell & Fazio, 1984). However, in contrast to work on repeated attitude expression, and also most previous work in self-persuasion, we find that advocates may paradoxically depolarize following a pro-attitudinal advocacy attempt.

It also may be argued that the lack of a control condition restricts the inferences which can be made about depolarization occurring during pro-attitudinal advocacy. We did not employ a control condition in the current work, as our aim was not to show whether self-persuasion exists or not (i.e. our aim was not to contrast pro-attitudinal advocacy with the absence of advocacy), but rather to investigate the processes occurring specifically when people advocate for their own opinions, as they do so frequently (see Briñol et al., 2012; Gordijn et al., 2001 for similar study designs). Further, we show that depolarization does not depend on audience attitudes (i.e. whether a target agrees, disagrees, or is neutral; see Supplementary Materials). This rules out alternative explanations for depolarization, such as, activation of knowledge supporting opposing views (see Förster & Liberman, 2007 for a review of knowledge activation).

Rather, depolarization in the current work is partially explained by meta-cognitive confidence in one's advocacy attempt, such that those who are less confident in their message are more likely to depolarize post-advocacy. This effect may be explained by the advocator's knowledgeability and previous experience advocating for their attitudes. People tend to experience difficulty justifying their own attitudes (Wilson & Dunn, 1986), and depolarization following justification tends to be moderated by knowledgeability of the issue, such that those who are less knowledgeable are more likely to depolarize (Wilson et al., 1989). Further, previous work in group polarization and counter-attitudinal advocacy show that those who are unfamiliar with the attitude issue tend to depolarize more (Flanders & Thistlethwaite, 1967; Janis & King, 1954). In our context of pro-attitudinal advocacy, advocators may have realized that they lack sufficient issue-specific knowledge to elaborate on their attitude position, realized shortcomings of their arguments, or both, thus experiencing low meta-cognitive confidence and resulting attitude depolarization.

Note, however, that we were unable to identify knowledgeability (measured as subjective source credibility) or any other *specific* meta-cognitive variable which can be successfully manipulated to induce depolarization. In previous self-persuasion research, meta-cognitions tend to be tested in isolation (thought confidence: Clarkson et al., 2011; 2013; perceived argument quality and perceived effort: Briñol et al., 2012), restricting the ability to test the separability of these variables. An exception is Briñol and colleagues (2012), who demonstrated that items measuring perceived argument quality and perceived effort were highly correlated (Cronbach's $\alpha = .85$) and thus aggregate across these items to form an index of argument quality. Similarly, our studies suggest some degree of overlap between these specific meta-cognitions, but also all other theoretically different meta-cognitions, for instance, those that are relevant to thoughts and not arguments, such as thought confidence and thought favorability (see Supplementary Materials for correlation

matrices). For example, thought confidence and argument quality were highly correlated in both our studies ($r_{\text{Study 1}} = .61$; $r_{\text{Study 2}} = .65$). Future research may hone in on the separability of these particular dimensions using reliable indicators designed to test construct distinctiveness.

Further, future work may examine which particular meta-cognitions impact self-persuasion to the greatest extent. In the current research, post-hoc regression analyses containing initial stance and all the meta-cognitive variables reveal perceived argument quality as the only unique meta-cognitive predictor of depolarization across both studies (Study 1: $\beta = .19$, $p = .02$; Study 2: $\beta = .15$, $p = .02$). Previous work shows that manipulating thought confidence via false feedback influences standard persuasion (Petty, Briñol, & Tormala, 2002; Tormala, Briñol, & Petty, 2004). In the context of self-persuasion, our results suggest that manipulating argument quality using a false feedback mechanism is more likely to produce depolarization compared to manipulating properties of thoughts. It appears that in order to promote openness to opposing viewpoints, practitioners should increase uncertainty by targeting specific meta-cognitions during pro-attitudinal advocacy, but also ensure that this happens under conditions of elaboration.

Moderating Role of NfC

The current research points to the moderating role of NfC on the links between meta-cognitive confidence, polarization, and intentions to engage with the opposition. We find that the relationships between meta-cognitive confidence and advocacy-related dependent variables are at least partly dependent upon the extent to which the advocator elaborates on his or her thoughts. While NfC leads to increased confidence on average, it moderates the effect of meta-cognitive confidence on intentions to engage with the opposition, though it does not moderate the effect of meta-cognitive confidence on proselytization. We find that when meta-cognitive confidence in advocacy is low, those who are high in NfC experience

greater depolarization and express increased willingness to interact with the opposition. This suggests an openness effect of NfC, such that, although NfC individuals are likely to be more confident in their advocacy attempt, in instances where they do in fact experience low confidence, they are more receptive to opposing viewpoints. Indeed, high NfC individuals are more likely than low NfC advocates to revise their beliefs if they believe that their attitude was based on biased reasoning processes (Wegener & Petty, 1997). Further, given that those in high NfC are more likely to deliberate on the validity of their own position (Briñol & Petty, 2003), and spend more time thinking about their thoughts (Petty et al., 2007), they are more likely than low NfC individuals to seek further information (e.g. from the opposition) to inform their attitude when confidence is lacking.

This increased desire to interact with dissimilar others is likely to be driven by motivation to hold more accurate views (i.e. accuracy motivation; see Chaiken & Trope, 1999). Given that high NfC individuals are more likely to revise their beliefs, and deliberate on the validity of their own position, they may be more attuned to *information utility* (Hart et al., 2009), or the benefits additional information has to offer. Those high in NfC may be more willing to seek attitude-inconsistent information in order to inform confidence in their own position, and thus be more willing to interact with dissimilar others. That is, the prospect of information utility may override confirmation bias (Knobloch-Westerwick & Kleinman, 2012) for individuals who are more willing and able to elaborate on the validity of their own position.

On the contrary, those who are low in NfC may experience the opposite pattern of results under conditions of low meta-cognitive confidence, possibly induced by a motivation to justify their own views (see Hart et al., 2009 for a review on defense versus accuracy motivations). That is, they may be more inclined to protect their own opinions (Festinger, 1962) when their confidence is threatened. This defense motivation drives advocates to

exhibit confirmation bias, thus leading them to seek attitude-confirming information or like-minded others (Nickerson, 1998), compared to seeking attitude-inconsistent information or dissimilar interaction partners. Future research may test the relationships between NfC and different types of motivations, to further disentangle the openness effect of NfC.

Limitations and Future Directions

The current research has a few key limitations. First, we focused on one specific attitude issue (carbon emissions policy). Thus, the generalizability of our findings to different issues should be investigated by future research. Different attitude issues likely vary in their valence, controversy, and their ability to evoke emotional reactions (Reeves, Yeykelis, & Cummings, 2016). Thus, we might expect depolarization and communicative intentions to differ based on the attitude issue tested. For example, attitudes towards controversial issues (e.g. capital punishment) are less likely to depolarize via self-persuasion, compared to others (e.g. product attitudes; see Vinokur & Burnstein, 1978). In the current research we replicate attitude depolarization across two studies for a relatively controversial issue relating to climate-change, which is promising. We also find depolarization to occur for people across the attitude spectrum, thus demonstrating that this effect occurs regardless of initial attitude position.

Second, it must be noted that correlation does not equal to causation. Although we find that meta-cognitive confidence predicts attitude depolarization and communicative intentions, future research is required to demonstrate a causal link. Similar to current work, even though past research typically manipulates an advocacy context, it tends to investigate the relationships between dependent variables in a correlational manner. Future work manipulating meta-cognitive confidence (e.g. via a false-feedback mechanism) is required to definitively establish meta-cognitive confidence as the driving factor behind depolarization, and not others (e.g. emotions). Nevertheless, our study provides important preliminary

evidence that meta-cognitive confidence is associated with depolarization, specifically in the context of pro-attitudinal advocacy.

Conclusions

The current research suggests that people may paradoxically depolarize even after advocating for their own position. This effect is driven by how much confidence people place in their advocacy attempt, such that those who experience low meta-cognitive confidence during advocacy are more likely to depolarize. Interestingly, this effect is moderated by one's tendency to elaborate, such that those who by disposition prefer complex thinking are more likely to depolarize and engage with those holding opposing views when meta-cognitive confidence is low. This finding suggests an important boundary condition to translating advocacy into open communication.

Supplementary Materials

Table 4-1

Spearman's Rho Correlations Between Meta-Cognitive Variables (Study 1)

	TF*TC	Thought confidence (TC)	Thought conviction	Thought favorabil ity (TF)	Argument quality	Perceived effort
Perceived effort	-.02	.23**	.19**	.00	.21**	
Argument quality	.18**	.61**	.43**	.21**		
Thought favorability	.97**	.19**	.18**			
Thought conviction	.17**	.49**				
Thought confidence	.13*					
TF*TC						

Note: **correlation is significant at the .01 level (two-tailed); *correlation is significant at the 0.05 level (two-tailed).

Table 4-2

*Summary of Multiple Regression Analyses for Variables Predicting Polarization in Study 1 – With TF*TC measure (N = 259)*

Variable	<i>B</i>	<i>SE B</i>	β
Initial stance	-.09	.08	-.07
Initial absolute attitude extremity	-.71	.13	-.37**
Time 1 attitude confidence	.18	.05	.22**
Meta-cognitive confidence	.15	.08	.12
Thought valence	-.11	.08	-.08
R^2			.14
$R^2_{Adjusted}$			.13
F			8.36

* $p < .05$. ** $p < .01$.

Table 4-3

*Summary of Multiple Regression Analyses for Variables Predicting Polarization in Study 1 – Without TC*TF measure (N = 259)*

Variable	<i>B</i>	<i>SE B</i>	β
Initial stance	-.06	.08	-.05
Initial absolute attitude extremity	-.73	.12	-.38**
Time 1 attitude confidence	.17	.05	.22**
Meta-cognitive confidence	.19	.08	.14*
R^2			.14
$R^2_{Adjusted}$			.13
F			10.38**

* $p < .05$. ** $p < .01$.

Effects of Target and Initial Stance (Study 2)

A 2(Initial stance: Pro vs Anti) x 3(Target: Agree, Disagree, Neutral) between-subjects ANOVA on polarization index showed no significant effect of target on depolarization $F(2, 489) = .66, p = .52$. However, unlike Study 1, an effect of initial stance emerged, $F(1, 489) = 7.83, p = .005, \eta p^2 = .02$. Pairwise comparisons revealed that the Anti advocates depolarized to a greater extent compared to Pro advocates ($M_{\text{Anti}} = -.46, SD = 1.25, M_{\text{Pro}} = -.19, SD = .94$). There was no interaction between target and initial stance, $F(2, 489) = 1.70, p = .19$.

In terms of open communicative practices, intentions to engage with the opposition was not affected by either target, $F(2, 489) = 2.21, p = .11$, nor initial stance, $F(1, 489) = .60, p = .44$). On the other hand, in terms of closed communicative practices, although intentions to persuade others was not affected by target, $F(2, 489) = 2.50, p = .78$, this was influenced by initial stance, $F(1, 489) = 18.00, p < .001; \eta p^2 = .04$, such that, Anti advocates were less likely to try and persuade others of their opinion ($M_{\text{Anti}} = 4.52, SE = .12, M_{\text{Pro}} = 5.21, SE = .11$). There was no interaction between target and initial stance, $F(2, 489) = .21, p = .81$.

Table 4-4

Spearman's Rho Correlations Between Meta-Cognitive Variables (Study 2)

	Source credibility	Thought confidence	Thought conviction	Thought favorability	Argument quality	Perceived effort
Perceived effort	.42**	.40**	.37**	.06	.57**	
Argument quality	.53**	.65**	.44**	.14**		
Thought favorability	.19**	.14**	.19**			
Thought conviction	.42**	.45**				
Thought confidence	.47*					
Source credibility						

Note: **correlation is significant at the .01 level (two-tailed); *correlation is significant at the 0.05 level (two-tailed).

Table 4-5

*Summary of Multiple Regression Analyses for Variables Predicting Polarization in Study 2
(N = 477)*

Variable	<i>B</i>	<i>SE B</i>	β
Initial stance	.13	.05	.12**
Initial absolute attitude extremity	-.41	.07	-.26**
Time 1 attitude confidence	.09	.03	.13**
Meta-cognitive confidence	.16	.05	.15*
R^2			.10
$R^2_{Adjusted}$			.09
F			13.37**

* $p < .05$. ** $p < .01$.

Table 4-6

Summary of Moderated Regression Analyses for Variables Predicting Polarization – Study 2
(N = 477)

Variable	<i>B</i>	<i>SE B</i>	LLCI	ULCI
Initial stance	.14**	.05	.01	.20
Initial absolute attitude extremity	-.42**	.07	-.57	-.27
Time 1 attitude confidence	.08**	.03	.02	.15
Meta-cognitive confidence (MCC)	.17**	.05	.07	.27
NfC	.02	.05	-.08	.12
NfC x MCC	.10*	.05	.01	.20
<i>R</i> ²				.11
<i>F</i>				9.80**

Note: * $p < .05$. ** $p < .01$.

LLCI = lower limit of 95% confidence interval for *B*.

ULCI = upper limit of 95% confidence interval for *B*.

Table 4-7

Summary of Moderated Regression Analyses for Variables Predicting Intentions to Engage – Study 2 (N = 477)

Variable	<i>B</i>	<i>SE B</i>	LLCI	ULCI
Initial stance	-.10	.07	-.24	.04
Initial absolute attitude extremity	.06	.11	-.15	.26
Time 1 attitude confidence	.04	.05	-.05	.13
Meta-cognitive confidence (MCC)	.15	.08	-.02	.31
NfC	.49**	.07	.35	.63
Intentions to persuade	.23**	.04	.14	.32
NfC x MCC	-.25**	.07	-.38	-.12
R^2				.23
F				20.35**

Note: * $p < .05$. ** $p < .01$.

LLCI = lower limit of 95% confidence interval for *B*.

ULCI = upper limit of 95% confidence interval for *B*.

Chapter 5 Moral vs Practical: Moral Expressiveness and Self-Persuasion

Preface to Chapter 5

(This chapter was previously published in the *Journal of Social and Political Psychology* as “Why Moral Advocacy Leads to Polarization and Proselytization: The Role of Self-Persuasion”; the accepted version of the article is included in this thesis)

The focus of Chapter 4 was to investigate the processes and consequences of pro-attitudinal advocacy. Specifically, we explored whether freely advocating one’s position leads to attitude depolarization, and if so, the mechanisms and boundary conditions of this effect. We found meta-cognitive confidence and Need-for-Cognition as key variables influencing not only attitude depolarization, but also communicative intentions, such as willingness to engage with dissimilar others.

In contrast, the key goals of this research stream were to: a) understand the implications of “constrained” (i.e. experimentally manipulated) as opposed to “free” pro-attitudinal advocacy on communicative intentions, and b) mechanisms other than for self-persuasion which may underly these effects. Thus, we examine the mechanisms and consequences of two distinct styles of advocacy: moral versus practical advocacy. Research in moral psychology and political psychology demonstrate the polarizing effects of advocacy grounded in moral rhetoric (Clifford, Jerit, Rainey, & Motyl, 2015; Tappin & McKay, 2019). For example, moral rhetoric has been linked with greater proselytization of one’s attitudes, the formation of echo chambers, and in extreme cases, hate crimes (Brady et al., 2017; Skitka & Mullen, 2002; Valenzuela et al., 2017). An alternative to moral reasoning that has been proposed as a potential solution to retaining free speech while minimizing social conflict, is

practical advocacy (see Kovacheff, Schwartz, Inbar, & Feinberg, 2018). In contrast to moral advocacy, practical advocacy encompasses reasoning grounded in cost-benefit analyses, and economic consequences, for example.

The distinction between moral and practical advocacy is important because these two types of advocacy may yield different attitudinal and behavioral outcomes. For example, research in standard persuasion demonstrates that arguments grounded in moral reasoning is more persuasive than those grounded in practical reasoning (Leidner et al., 2018). Despite these potential differences, research in advocacy and persuasion is yet to empirically contrast moral and practical advocacy. Thus, the novelty of the current research is in: a) exploring an alternative form of reasoning to moral advocacy (i.e. practical advocacy) in the context of self-generated persuasion, and b) identifying the ways in which these two types of advocacy exert their effects on polarization and proselytization.

Given that a variety of appeals may be used to advocate one's attitudes, we examine four distinct mechanisms by which moral versus practical advocacy may either exacerbate or attenuate polarization and proselytization: a) expression of moral foundational values (harm, fairness, loyalty, authority, purity; see Graham et al., 2011), b) reliance on moral systems (deontological reasoning based on rules, versus consequentialist reasoning based on outcomes; see Wheeler & Laham, 2016), c) expression of emotive language, specifically emotions which signify moral outrage (i.e. anger and disgust; see Haidt, 2003), and finally d) confidence in one's advocacy attempt (i.e. meta-cognitive confidence; see Petty & Briñol, 2015 for a review). Overall, our findings suggest that practical, compared to moral advocacy is likely to attenuate polarization and proselytization, primarily via one of the four proposed mechanisms.

Citation: **Abeywickrama, R. S.**, Rhee, J. J., Crone, D. L., & Laham, S. M. (2020). Why

Moral Advocacy Leads to Polarization and Proselytization: The Role of Self-Persuasion.

Journal of Social and Political Psychology, 8(2), 473–503.

<https://doi.org/https://doi.org/10.5964/jspp.v8i2.1346>

**Why Moral Advocacy Leads to Polarization
and Proselytization: The Role of Self-
Persuasion**

Abstract

This research is the first to examine the effects of moral versus practical pro-attitudinal advocacy in the context of self-persuasion. We validate a novel advocacy paradigm aimed at uncovering why moral advocacy leads to polarization and proselytization. We investigate four distinct possibilities: (1) expression of moral foundational values (harm, fairness, loyalty, authority, purity), (2) reliance on moral systems (deontology and consequentialism), (3) expression of moral outrage, and (4) increased confidence in one's advocacy attempt. In Study 1 ($N = 255$) we find differences between moral and practical advocacy on the five moral foundations, deontology, and moral outrage. In Study 2 ($N = 218$) we replicate these differences, but find that only the expression of moral foundations is consequential in predicting attitude polarization. In Study 3 ($N = 115$) we replicate the effect of moral foundations on proselytization. Our findings suggest that practical compared to moral advocacy may attenuate polarization and proselytization. This carries implications for how advocacy can be re-framed in ways which minimize social conflict.

Keywords: polarization, morality, persuasion, advocacy, migration, moral foundations, self-persuasion, attitudes

Non-technical summary

Background

Society is becoming increasingly polarized, and conflict between groups holding opposing views is frequent. A potential source of this conflict is people frequently advocating for their opinions in terms of moral values, which tend to be perceived as universal and sacred. However, we do not know the specific consequences of moral advocacy, and how it can lead to negative outcomes (e.g. political polarization). We also do not know if there are other forms of advocacy which may attenuate these negative outcomes.

Why Was This Study Done?

In light of the currently fractured and polarized attitudinal landscape, in which advocacy is frequent, we sought to understand: (1) the consequences of moral advocacy on attitude polarization (attitudes becoming more extreme), and proselytization (people's willingness to persuade others of their own opinion), (2) the specific kinds of moral values or moral content which lead to such negative outcomes, and (3) alternative ways in which advocacy might be encouraged (e.g. advocacy grounded in "practical" or economic arguments as opposed to moral arguments) to reduce the likelihood of such outcomes.

What Did the Researchers Do and Find?

We conducted an online survey in which 588 people were asked to write arguments for their position on migration and climate-change issues. People were randomly divided to advocate in moral or practical terms. We measured their attitudes pre- and post-advocacy, and the confidence they placed in their advocacy attempt. We also calculated scores on the kinds of moral content expressed in their arguments, for example, the extent to which they expressed moral values such as harm, care, purity, and emotions such as anger and disgust. Our findings showed that moral (versus practical) advocacy leads to people's attitudes

becoming more extreme post-advocacy, and increased willingness to persuade others of one's opinion. This was due to increased expression of moral values, as opposed to other types of moral content, such as emotions, or confidence in advocating.

What Do These Findings Mean?

Our findings suggest that re-framing advocacy away from moral values, and towards more “objective” practical arguments such as, economic consequences, leads to less polarization and proselytization. This means that by encouraging practical advocacy, we might be able to minimize social conflict, while simultaneously retaining our right to free speech.

Introduction

The US federal government shutdown in response to the erection of a border wall around Mexico reflects the failure of society to make compromises on morally-charged issues (Delton, DeScioli, & Ryan, 2019; Ryan, 2017). The current political climate is undoubtedly one of disagreement and polarization, in which conflict between ideologically opposing groups is frequent. Indeed, partisan antipathy in the United States is at its highest level since 1994, with both Republicans and Democrats now being more likely to report unfavorable views of the other party, and even perceive the other party's policies as a threat to the nation's well-being (Pew Research Center, 2014). An often-cited cause of this social division is the increased prevalence of moral rhetoric as a foundation for political advocacy (Clifford et al., 2015; Tappin & McKay, 2019).

Despite the destructive effects of moral rhetoric, little is known about the mechanisms by which moral rhetoric leads to increased social division. Specifically, it is currently not known how relying on moral rhetoric during pro-attitudinal advocacy (advocating for one's own position; see Briñol, McCaslin, & Petty, 2012; Gordijn, Postmes, & de Vries, 2001), changes the advocate's *own* attitudes and communicative intentions. We propose that relying on moral rhetoric (compared to non-moral rhetoric) during pro-attitudinal advocacy leads to greater attitude polarization, and increased proselytization (persuading others of one's opinions), via *self-persuasion* (see Briñol et al., 2012; Tesser, 1978; Wilson, 1990).

Importantly, the current research explores the mechanisms through which pro-attitudinal advocacy leads to polarization and proselytization with the aim of uncovering how advocacy can be re-framed to reduce closed communicative practices. For this purpose, the current research contrasts "moral" versus "practical" expression of one's attitudes in the pursuit of: a) validating a novel advocacy task aimed at changing the language people use to

advocate for their views, b) exploring how the extent to which people rely on moral advocacy may polarize one's attitude post-advocacy, or increase one's desire to proselytize one's attitudes, and c) exploring four possible mechanisms – moral value framing (individualizing and binding moral foundations), moral system framing (deontology vs consequentialist reasoning), emotional language use (moral outrage), and meta-cognitive confidence, which may account for polarization and proselytization.

The Problem with Moral Rhetoric

One potential source of political polarization is the introduction of moral rhetoric to justify attitudes. Extant research in moral psychology suggests that when an issue is construed as moral, it evokes moral emotions such as disgust or anger (Haidt, 2003; Tangney, Stuewig, & Mashek, 2007), and comes to be perceived as universal (Skitka, Bauman, & Sargis, 2005; van Bavel, Packer, Haas, & Cunningham, 2012) and beyond compromise (for a recent review see Rhee, Schein, & Bastian, 2019; Ryan, 2017; Skitka, 2010). Such changes in the construal of the moralized attitude affects interpersonal interaction, often causing individuals to try to distance themselves from alternative viewpoints (Frimer, Brandt, Melton, & Motyl, 2019; Wright, Cullum, & Schwab, 2008), and even perceive opinions on the relevant idea in question as more polarized (Anderson et al., 2014; Ryan, 2017), which can in turn lead to erosion of social trust (Rapp, 2016).

This work in moral psychology is consistent with work in standard persuasion (i.e. an external source delivering a persuasive message to a recipient; see Petty & Briñol, 2014 for a recent review), which suggests that moral framing increases polarization (Clifford, 2019; Clifford et al., 2015). For example, Clifford (2019) found that presenting participants with persuasive messages framed in moral terms (e.g. harm) led participants to state that they would be more upset if people close to them disagreed with their position on the issue.

In addition to polarization, research has also demonstrated that moral rhetoric is a strong motivator of proselytization. For example, in a study of articles published in six Chilean outlets, Valenzuela, Piña, and Ramírez (2017) found that news stories were more likely to be shared on social media when they were framed in moral terms, while they were less likely to be shared if framed in terms of economic consequences or conflict. Although morally framed issues are more likely to be shared among one's own ingroup, it has also been found that morally emotive rhetoric is less likely to be shared with outgroup members, leading to the formation of attitudinal "echo-chambers" (Brady, Wills, Jost, Tucker, & Van Bavel, 2017).

Thus, while moral framing of attitudes may be particularly effective in disseminating an attitude among like-minded others, and even mobilizing in-group members to advance a group's position (van Zomeren, 2013; van Zomeren, Postmes, & Spears, 2012; van Zomeren, Postmes, Spears, & Bettache, 2011) it can also be a roadblock to engagement in constructive discussion and compromise with those who hold opposing views.

Pro-Attitudinal Advocacy and Self-Persuasion

Despite research demonstrating the consequences of moral rhetoric on polarization and proselytization, less is understood about the mechanisms by which this occurs. To address this gap in understanding, we focus on the relationships between moral rhetoric, polarization, and proselytization, in the context of *self-generated* persuasion. Given that highly convicted attitudes tend to be difficult to change using standard persuasion (Krosnick & Petty, 1995), an alternative to standard persuasion that has been used over the last few decades (although not examined as frequently), is self-persuasion. While in standard persuasion the source and recipient are two separate entities, in self-persuasion, the message is generated and received by the same person (Briñol et al., 2012; Hovland et al., 1953; Wilson, 1990). Thus, while standard persuasion examines the consequences of an external

persuasive message on a third-party recipient, self-persuasion examines the consequences of recipients generating persuasive messages *themselves*. This implies that when one advocates for one's position on a topic (i.e. pro-attitudinal advocacy), one's attitudes and communicative intentions may change due to self-persuasion. For example, one may become more polarized or entrenched in one's opinion towards migration, after advocating for one's position on this topic. We examine the mechanisms and consequences of moral versus practically-framed advocacy appeals in the context of pro-attitudinal advocacy, given its high ecological validity and prevalence in the current political climate.

To our knowledge, this is the first set of studies to explore the effects of moral versus practical advocacy in the context of self-persuasion occurring during pro-attitudinal advocacy. We sought to understand if: a) people can self-persuade to take a more or less extreme position (polarize or depolarize) after advocating for their opinion, and b) even if attitudes do not change, how advocacy may inform subsequent communicative intentions, such as intentions to proselytize.

The second gap in the literature we aim to address is creating a specific advocacy paradigm (i.e. moral versus practical advocacy) that can be used in the context of self-persuasion to depolarize attitudes and reduce closed communicative intentions, such as intentions to proselytize. Previous research has shown effects of moral framing on attitude polarization, whereby moral framing leads to greater polarization and unwillingness to compromise (Clifford, 2019; Ryan, 2017). Although a growing body of literature now suggests that the use of moral language in rhetoric can lead to polarization, few studies have investigated methods of persuasion that may depolarize an already polarized attitude (see Feinberg & Willer, 2013 for an exception). Although many have proposed rational argumentation or non-moral framing as a potential solution to reducing polarization (Kovacheff et al., 2018), this is yet to be tested empirically in the context of advocacy. We

suggest that one way to reduce polarization and encourage more open communication between ideological groups, is to change the way in which people communicate their own attitudes and beliefs. That is, we suggest that reframing one's attitudes in "practical" as opposed to moral terms is less likely to lead to depolarization, and lower intentions to proselytize one's attitudes.

Proposed Alternative: Practical Reasoning

Given the potential of moral rhetoric to stimulate social conflict, we explore an alternative method one can use to justify one's attitudes. We term this alternative "practical" reasoning. This type of reasoning is based on economic consequences and cost-benefit analyses. It is commonly described by economic theory, whereby the best course of action is the one that allocates scarce resources to the option that "maximizes expected utility" (Kahneman & Tversky, 1986; Simon, 1979; Tversky & Kahneman, 1981). Compared to moral reasoning, practical reasoning is arguably a more withdrawn, less convicted and more "objective" or value-free way of expressing one's attitudes.

The moral versus practical distinction is important, given that attitudes and arguments grounded in these two kinds of reasoning tend to yield different attitudinal outcomes (Leidner et al., 2018; Luttrell, Petty, Briñol, & Wagner, 2016; Luttrell, Phillip-Muller, & Petty, 2019; Wheeler & Laham, 2016). For example, research intersecting standard persuasion and political psychology shows that presenting arguments grounded in moral values as opposed to those grounded in cost-benefit analyses are more effective in reducing support towards torture in the US (Leidner et al., 2018). Further, research in social psychology examining the consequences of attitude properties, shows that attitudes that are based on morality are more consequential in predicting behavior than those based on non-moral reasoning (Luttrell et al., 2016).

Morally framed messages and attitudes tend to be more persuasive, and likely lead to greater polarization and proselytization compared to non-moral attitudes, in the context of standard persuasion. We test whether this effect generalizes to the context of self-persuasion, and predict that the effect of moral advocacy is driven by the saliency of moral content contained in moral compared to non-moral messages. Unlike previous work examining the effects of attitude bases, or the effects of presenting moral versus non-moral messages, the focus of the current study is on moral versus *practical* message content generated by the advocate *themselves* during pro-attitudinal advocacy. To our knowledge, this is the first study contrasting these two types of message content in the context of self-persuasion. We predict that using moral arguments instead of practical arguments to justify one's attitudes will lead to greater polarization and increased intentions to proselytize one's attitudes, via four possible mechanisms: moral values (moral foundations), moral systems framing (deontology vs consequentialism), moral emotions (moral outrage), or confidence in one's advocacy attempt (meta-cognitive confidence).

Proposed Mechanisms

People use a variety of frames and appeals to both justify and promote their attitudes. A relatively common way in which individuals express their attitudes is through the use of moral rhetoric – that is, framing attitudes in terms of moral values (Luttrell et al., 2019). Another way of expressing one's attitudes is by relying on moral systems, that is, deontological reasoning (based on rules and principles) and consequentialist reasoning (based on outcomes or consequences; see Wheeler & Laham, 2016). Yet another possibility is relying on emotive language, specifically, “other-condemning” emotions such as anger and disgust, which signify moral outrage (see Haidt, 2003). Another possibility that does not capture the language of advocacy, but rather, the confidence placed on one's advocacy

attempt, is *meta-cognitive confidence* (see Petty & Briñol, 2015 for a review). We discuss each of these four possibilities in turn.

Moral Foundations Theory

Over the last two decades, the field of moral psychology has made substantial advances in the conceptualization and study of mechanisms underlying moral decision making and moral values (Greene, 2015; Haidt, 2007). A powerful driver of such advances has been the development of new theoretical frameworks which seek to define psychological domains that underlie morality-specific cognition and decision-making (Haidt, 2007). Perhaps the most prominent of these accounts has been Moral Foundations Theory (MFT: Haidt & Graham, 2007) which posits that there are five fundamental domains of moral values which represent innate sources of moral intuitions present in all individuals – namely, harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity. A key tenet of MFT is that, although these five domains of moral intuition are present in all individuals, there may be substantial differences between societies, and subgroups within societies, in the degree to which a given domain may be emphasized or elaborated upon (Graham, Haidt, & Nosek, 2009; Haidt & Graham, 2007; Koleva, Graham, Iyer, Ditto, & Haidt, 2012).

Moreover, MFT proposes that political liberals place greater emphasis on harm/care and fairness/reciprocity (collectively the individualizing foundations), while political conservatives place emphasis on ingroup/loyalty, authority/respect, and purity/sanctity foundations (the binding foundations; Graham et al., 2009; Haidt & Graham, 2007). For simplicity, we refer to the combination of these two foundation clusters as “moral expressiveness”, which captures the extent to which one relies on both the individualizing and binding foundations to justify one’s attitude.

Within the context of moral psychology research, MFT approaches have demonstrated particular utility in explaining political, ideological and cultural divides (Graham et al., 2009; Haidt & Graham, 2007). For example, Koleva et al. (2012) found that the degree to which individuals endorse certain moral foundations were strongly predictive of their position on a number of highly polarizing political issues such as same-sex marriage, euthanasia, and cloning. Indeed, many of these positions were more strongly predicted by the level of endorsement of certain foundations (especially purity/sanctity) than by political orientation or religious belief.

The finding that highly polarized positions on political issues tend to coincide with strong endorsement of certain moral foundations raises the possibility that moral foundations endorsement may in fact be an underlying mechanism for polarization. Supporting this idea, Mooijman, Hoover, Lin, Ji, and Dehghani (2018) found that the frequency of moral foundations terms appearing in Tweets sent during the 2015 Baltimore protests (Yan & Ford, 2015) could predict both hourly police arrest numbers in the Baltimore area, and whether the following day was likely to feature a violent or peaceful protest. This suggests that exposure to rhetoric containing moral foundations language led to more extreme endorsement of a given position. In the context of self-persuasion, this also suggests that moral, compared to practical advocacy is likely to lead to greater attitude polarization, and increased willingness to proselytize one's opinions, depending on the extent to which one relies on moral values to justify one's position.

Deontology vs Consequentialism

Research has also found that people may strategically or intuitively appeal to different types of moral systems when seeking to justify their positions related to moral issues. For example, Piazza and Sousa (2014) suggest that people are more likely to rely on *deontological* appeals (duties, principles or rules about right or wrong behavior), than

emotive (elicitation of emotional reactions) or consequentialist reasoning (outcomes, consequences, or effects) in making moral judgements. Further, using content-based analyses, Wheeler and Laham (2016) found that participants were more likely to appeal to consequentialist reasoning when asked to justify their position on issues relevant to the individualizing foundations than the binding foundations, for example. These findings suggest that moral advocacy is more likely to encourage advocates to use words related to deontology (and potentially consequentialism) compared to practical advocacy.

Moral Emotions

Another possibility that we consider, is that moral advocacy leads to increased polarization and proselytization via greater expression of moral emotions. While moral messages tend to be perceived as more emotional compared to practical messages, in general (Brady et al., 2017; Luttrell et al., 2019), we specifically consider emotions conveying moral outrage such as anger and disgust (Haidt, 2003; Salerno & Peter-Hagene, 2013). Moral outrage refers to emotions that are expressed in response to a perceived moral norm violation (Crockett, 2017; Haidt, 2003; Salerno & Peter-Hagene, 2013). Extant research suggests that people may be particularly motivated to share their expressions of moral outrage via gossip, shaming and punishment as a means of signaling their moral character to other ingroup members (Jordan, Hoffman, Bloom, & Rand, 2016), and reaffirming ingroup norms (Crockett, 2017).

Consistent with this notion, recent research conducted by Brady et al. (2017), found that posts on Twitter regarding a morally laden issue (e.g. same-sex marriage, gun control, climate-change) were more likely to be retweeted when they contained words expressing moral outrage. It was also found that this effect of moral-emotion words on increased sharing was amplified among networks of individuals who shared the same ideological position as the initial Tweet. Taken together, such findings suggest that receiving moral rhetoric

containing expressions of moral outrage may increase proselytization, and also increase polarization by promoting increased message sharing within ideological echo chambers.

Meta-Cognitive Confidence

A final possibility is that moral versus practical advocacy generates different levels of *meta-cognitive confidence*. Meta-cognitive confidence captures the confidence one places in one's thoughts and judgements (see Wagner, Briñol, & Petty, 2012 for a review). For example, one may perceive one's arguments to be strong or weak (Briñol et al., 2012), perceive oneself to have expended high or low effort in generating arguments (Briñol et al., 2012), or perceive oneself to be a knowledgeable or an unknowledgeable source on the topic (Ehret et al., 2018; Rios et al., 2018). Meta-cognitive confidence is important because it has been shown to be a key driver of polarization not only in the context of self-persuasion more generally (see Clarkson, Tormala, & Leone, 2011; Clarkson, Valente, Leone, & Tormala, 2013), but more specifically during self-persuasion occurring as a result of pro-attitudinal advocacy. Specifically, increasing meta-cognitive confidence, as indexed by perceived argument quality and effort, has been shown to lead to greater polarization following pro-attitudinal advocacy (Briñol et al., 2012).

Moreover, research from standard persuasion provides evidence for a “moral-matching” effect, such that, those who possess attitudes grounded in moral concerns are more likely to be persuaded by moral messages (Luttrell et al., 2019). This suggests that attitude basis may moderate the extent to which we observe polarization, due to match-induced processing fluency (Mayer & Tormala, 2010). That is, matching effects may increase the ease of message processing, increase meta-cognitive confidence and thus polarization post-advocacy.

Drawing on this literature, we hypothesize that asking one to advocate for one's position using moral as opposed to practical concerns, likely leads to increased meta-

cognitive confidence in one's attempt. Given the emphasis placed on the universality and correctness on one's own moral values, advocates are more likely to experience processing ease and fluency when justifying their position in moral terms. Moral values are highly sacralized and ingrained in our minds, and we use these as a basis to view the world (Kovacheff et al., 2018). People tend to think that their own moral values are both factual and universal (Tetlock, 2003), and thus may have difficulty understanding worldviews which deviate from one's own (Feinberg & Willer, 2013, 2015). Therefore, justifying one's position using moral reasons likely leads to increased meta-cognitive confidence, which may translate into attitude polarization, and/or greater willingness to persuade others of one's own opinion.

The Current Research

The main aim of this research is to explore whether moral versus practical advocacy changes levels of moral language and advocacy-related confidence during pro-attitudinal advocacy related to two compelling socio-political issues: migration and climate-change. One possibility is that moral advocacy increases grounding in foundational moral values. Another is that moral advocacy increases grounding in broader moral frameworks (e.g. deontology vs. consequentialism). A third possibility is that moral advocacy increases emotional grounding, or expression of moral outrage. Yet another possibility is that moral advocacy simply increases confidence in one's advocacy attempt. Thus, we have four possible accounts of the influence of moral advocacy on polarization, which we explicitly test using a new paradigm and text-based analysis.

In Study 1, we consider the extent to which moral versus practical advocacy influences: (a) foundational value expression, (b) moral framework language, and (c) moral outrage. In Study 2, we consider the extent to which these factors account for any effects of framing on polarization, by additionally examining the effects of meta-cognitive confidence. In Study 3, we extend our dependent variables to include communicative practices, such as,

intentions to proselytize one's opinions. Across the three studies, we seek to explore which of the four possible mechanisms likely accounts for the effects of moral advocacy on social division.

Study 1

Method

Participants and design. Two-hundred and fifty-five US residents were recruited via Amazon Mechanical Turk ($M = 40.71$, $SD = 12.45$, female = 138). An a-priori power analysis using G*Power 3.1 revealed that this sample size was sufficient to detect small-medium effects ($f = .20$) typically found in social psychology (Richard et al., 2003), with 80% power in an ANCOVA (Condition: Moral vs Practical) including two covariates (initial stance, political orientation) (Faul, Erdfelder, Lang, & Buchner, 2009).

Materials and procedure. Participants completed a survey online via Qualtrics (Qualtrics, 2016). Participants completed an attitude questionnaire assessing favorability towards various issues (e.g. carbon emissions, consumer behavior). Our issue of interest was one relating to migration (see Supplementary Materials, Appendix A for detailed attitude issue description).

Initial stance. Participants were asked to indicate their attitude towards migration using the following item, "Compared to the number of migrants the government usually takes, how many migrants do you think the US government should take?" on a 7-point scale (1 = *much fewer*, 7 = *a lot more*, $M = 4.07$, $SD = 1.79$). Those who indicated scale-point 1-3 were classified as initial stance = Anti (coded -1; $n = 77$), 4 = neutral (coded 0, $n = 78$) and 5-7 as initial stance = Pro (coded 1, $n = 100$).

Following initial attitude measurement, participants were randomly allocated to write advocacy appeals grounded in either moral ($n = 123$), or practical concerns ($n = 132$). The instructions for each appeal type can be found in the Supplementary Materials (Appendix B).

Demographics. Finally, participants were presented with a few demographic questions (age, gender, political orientation). Our primary measure of political orientation was a single 7-point Likert scale (1 = *very liberal*, 7 = *very conservative*; $M = 3.69$, $SD = 1.86$ (adapted from Feinberg & Willer, 2015; van Leeuwen & Park, 2009). A secondary political orientation measure was party affiliation with four choice options (Republican party = 34%, Democratic party = 50%, Libertarian party = 4%, other = 12%).

Characterizing moral content. Traditionally, moral content analyses have often been facilitated by word frequency-based approaches such as the Moral Foundations Dictionary (MFD: Graham et al., 2009) which identifies a number of keywords that are representative of each of the five foundations. Using moral foundations theory as a criterion for the presence of moral content in text can be a particularly powerful tool, as it allows moral psychology research to be extended into the study of freely generated text, which is common in online social-media platforms such as Twitter. Indeed, this method has been found to be highly effective in identifying moral content (Sagi & Dehghani, 2014a, 2014b), and also in predicting real-world moral behavior (e.g. Mooijman et al., 2018).

Thus, to measure the moral content of participants' text responses, we used distributed dictionary representations (Garten et al., 2018) covering different kinds of moral content (described below; see Garten, Boghrati, Hoover, Johnson, & Dehghani, 2016; Sagi & Dehghani, 2014a, 2014b for more examples). Distributed dictionaries perform a similar function to the more commonly-used raw frequency-based content analysis approach implemented in LIWC (Tausczik & Pennebaker, 2009), in that they measure the presence of pre-defined semantic content domains in a set of texts. However, distributed dictionaries differ from LIWC analyses in that they do not rely on raw word counts (i.e., the presence of keywords in texts). Rather, they rely on the overall semantic similarity of texts to the provided keywords, providing an arguably more sensitive measure of semantic content.

Specifically, distributed dictionaries measure semantic content using a vector space model, derived from a large training corpus, that represents words as points in multidimensional space, such that words with similar meanings (e.g., chair, seat) are closer together in that semantic space (i.e., have similar values on the model dimensions) by virtue of the fact that those words tend to occur in similar linguistic contexts in the training corpus. As in Garten et al. (2018), we operationalized semantic similarity as the cosine similarity (ranging from -1 to 1) of two vectors representing: (1) a given participants' text response, obtained by computing the average vector for all the words in the response, and (2) a dictionary, obtained by computing the average vector for all the words in a given dictionary category (e.g., the negative pole of the care foundation)¹⁷.

We created dictionary-based measures for three different kinds of moral content as follows. First, we measured semantic similarity of participants' responses to each of the five moral foundations, using dictionaries from previous research, where the positive and negative poles of each moral foundation are represented by four keywords (e.g., the words *suffer*, *cruel*, *hurt* and *harm* representing the negative pole of the care foundation; Garten et al., 2018; Hoover, Johnson, Boghrati, Graham, & Dehghani, 2018). Second, we developed additional moral content dictionaries based on the research of Wheeler and Laham (2016), covering specific moral framings (consequentialist and deontological), and third, capturing the expression of three different moral emotions relating to moral outrage (contempt, anger, disgust). This served to test the kind of moral language which primarily drives the effects of moral versus practical advocacy on polarization and proselytization (see Supplementary Materials, Appendix C for more information on how the dictionaries were constructed).

¹⁷ Dictionaries were constructed using the *embedding tools* package for R (Crone, 2018), using a pre-trained GloVe model (glove.42B.300d; Pennington, Socher, & Manning, 2014), limiting the model vocabulary to the 250,000 most frequently occurring words.

Indexing Moral Expressiveness

Moral expressiveness was conceptualized using the two broad categories of the five moral foundations: individualizing and binding foundations. A composite score on the individualizing foundations was created by averaging across z-scores of the following dictionary categories: harm, care, fairness, and cheating. A composite score on the binding foundations was created by averaging across z-scores of loyalty, betrayal, authority, subversion, purity and degradation. The scores on these measures represent semantic similarity of the text to the concept, such that a score of 0 means average similarity to the concept, negative scores imply relative dissimilarity, and positive scores indicate greater similarity.

Results and Discussion

Participants' arguments were screened for being on topic and adherence to the manipulation – no participants were excluded on this basis.

Differences between conditions. A principal components analysis was conducted to determine if the three moral emotions theorized (anger, contempt, disgust) load onto one component (i.e. a moral outrage component). The analysis revealed that all three moral emotions loaded onto a single component explaining 87% of the variance. This component represents *moral outrage* (component loadings: anger = .90, contempt = .94, disgust = .96), such that higher scores reflect greater moral outrage following one's advocacy attempt.

In order to assess differences between conditions on levels of moral outrage, moral expressiveness, and moral systems framing, a series of One-Way ANCOVAs were conducted with one between-subjects factor (Condition: Moral vs Practical) with initial stance and political orientation as covariates. Significant omnibus tests were followed by Bonferroni-corrected tests of multiple comparisons. Results revealed that our advocacy manipulation was successful - the moral and practical conditions significantly differed on most moral content tested (see Table 5-1 below). Almost all significant differences represent medium to very

large effects according to classification of effect sizes (η_p^2 small = .01, medium = .06, large = .14; Cohen, 2013), indicating substantial differences in the language used across the two advocacy conditions. Correlations between all variables are included in Table 5-2 below.

Table 5-1

Means and SEs for all dictionary scores as a function of Condition (Moral versus Practical), controlling for initial stance and political orientation in Study 1

	Moral	Practical	<i>F</i>	Partial η^2
	<i>M (SE)</i>	<i>M (SE)</i>	(<i>df</i> = 251)	
Moral foundations				
Individualizing	.29 (.07)	-.27 (.06)	37.59**	.13
Binding	.39 (.06)	-.37 (.07)	73.76**	.23
Deontology	.25 (.08)	-.24 (.08)	16.73**	.06
Consequentialism	-.08 (.09)	.08 (.09)	01.67	.01
Moral outrage	.45(.08)	-.42(.08)	58.72**	.19

Note: * $p < .05$, ** $p < .001$, two-tailed tests.

Table 5-2

Pearson's Correlations Between Continuous Variables in Study 1

	IS	PO	MO	CONS	DNT	BIND	IND
IND	-.03	-.14*	.52*	.35**	.80**	.87**	
BIND	.00	-.13*	.52**	.15*	.77**		
DNT	-.01	-.14*	.12*	.55**			
CONS	.04	.02	-.25**				
MO	.01	.03					
PO	-.41**						
IS							

Note: **correlation is significant at the .01 level (two-tailed); *correlation is significant at the 0.05 level (two-tailed). Shorthand notation: Individualizing foundations = IND; Binding foundations = BIND; Deontology = DNT; Consequentialism = CONS; Moral outrage = MO, Political orientation = PO, Initial stance = IS.

The results of Study 1 demonstrate that explicitly requesting advocates to frame pro-attitudinal advocacy in practical versus moral language changes the moral content of their appeal. Using a novel advocacy paradigm, we show that encouraging advocates to frame their appeal in moral versus practical terms increases the level of moral expressiveness, moral outrage, and deontological framing used to construct one's appeal relating to migration attitudes¹⁸. Despite differences in average levels of moral content, we do not know if these differences are consequential in predicting attitudes or communicative intentions post-advocacy.

¹⁸ We conducted a few additional analyses to test if our effects were driven by demographic characteristics, such as, political orientation. A linear regression containing initial stance, political orientation, Condition, and the individualizing (binding) foundations as independent variables revealed that political orientation was not a significant predictor of binding (individualizing) foundations ($\beta = -.04, p = .31$; $\beta = -.02, p = .46$). Given previous work on moral reframing in the standard persuasion literature, we might have expected political orientation to predict moral expressiveness, such that increased political conservativeness predicts increased expression of the binding foundations, while increased liberalism predicts greater expression of the individualizing foundations. However, this was not the case in the context of self-persuasion. There was also no difference in foundation scores between those who endorsed the different political parties, as captured by our secondary measure of political orientation. As part of a larger study we additionally measured personality variables such as Need-for-Cognition and the Big 5, however these were not analyzed as moderation by personality factors was not part of the central research question.

We attempt to replicate our findings in Study 2 with a few key modifications: a) using a different controversial attitude issue (i.e. issue relating to climate-change), b) a different sample population (college students versus MTurk population), c) including measures of pre and post-advocacy attitudes to assess attitude polarization as a result of advocacy, and d) to examine which of the following mechanisms: moral expressiveness, moral systems framing, moral emotions, or meta-cognitive confidence drives attitude polarization.

Study 2

Method

Participants and design. Two-hundred and fifty-five psychology undergraduates from an Australian university were recruited in exchange for course credit ($M = 19.34$, $SD = 2.10$, female = 164, other = 3). While two-hundred participants would have been sufficient to detect small-medium effects ($f = .20$) with 80% power in an ANCOVA (Condition: Moral vs Practical) with two covariates (initial stance, political orientation), we oversampled to allow for potential exclusions. The sample size was also sufficient to detect small-medium effects ($r = .20$) in two-tailed correlational analyses with 80% power (Faul et al., 2009).

Materials and procedure.

Similar to Study 1 participants completed the study on Qualtrics (2016). Participants were first asked to indicate their Time 1 attitudes and attitude basis.

Initial stance. Participants were presented with information on a hypothetical carbon emissions policy which was modelled based on energy-saving recommendations in Australia (“Department of Environment and Energy,” 2018). Participants were asked how favorable they felt towards a policy specifying a maximum thermostat temperature of 20 degrees Celsius during winter on a 7-point Likert scale (1 = *extremely unfavorable*, 7 = *extremely favorable*, $M = 4.18$, $SD = 1.79$; see Supplementary Materials, Appendix A for full description of attitude issue). Neutral participants (scale point = 4, $n = 37$) were excluded from the study, leaving a final sample of $N = 218$ (Anti = 97, Pro = 121). This is because attitude (de)polarization cannot be conceptualized and calculated clearly for those who do not indicate an initial preference (we elaborate on the calculation of attitude depolarization scores below).

Attitude basis. Two items were then used to assess the extent to which participants’ attitudes were based on moral and practical concerns (“To what extent is your attitude

towards migration based on moral [practical] concerns?”, 1 = *not at all based*, 7 = *very much based*; adapted from Teeny and Petty (2018); moral concerns $M = 4.33$, $SD = 1.74$; practical concerns $M = 5.21$, $SD = 1.49$). This was to see if the “moral-matching effect” found in standard persuasion generalizes to the context of self-persuasion (although not a key hypothesis), whereby moral messages should be more persuasive for those whose attitudes are based on moral concerns (see Luttrell et al., 2019).

Advocacy task. Participants were randomly allocated to either moral ($n = 108$) or practical advocacy conditions ($n = 110$) and presented with the same instructions as Study 1 with minor re-wording to account for the new attitude issue.

Thought-listing. Participants listed up to eight thoughts they had in response to generating their arguments (taken from Clark et al., 2013) before listing their post-advocacy attitude.

Time 2 attitude. Participants indicated their Time 2 attitude towards the carbon emissions policy ($M = 4.46$, $SD = 1.82$). Next, participants completed the following meta-cognitive measures related to their arguments in a randomized order: perceived effort, perceived argument quality and source credibility.

Perceived effort. Participants indicated how much effort they expended in generating their arguments on three 9-point scales (e.g. “How much energy did you put into generating your arguments, taken from (Briñol et al., 2012). Responses were averaged to create a composite measure of perceived effort ($M = 5.58$, $SD = 1.41$, Cronbach’s $\alpha = .84$).

Perceived argument quality.

Participants rated how strong they perceived they own arguments to be on three 9-point scales (e.g. “How strong are the arguments that you generated?”; adapted from Briñol et al., 2012). Responses were averaged to create a reliable composite score ($M = 5.20$, $SD = 1.67$, Cronbach’s $\alpha = .88$).

Self-perceived knowledgeability.

Participant knowledgeability on the carbon emissions issue was assessed using four 9-point scales (“How knowledgeable do you feel about energy policies?”, “How much information do you feel you have about energy policies?”, “To what extent do you feel you have expertise on energy policies?”, “To what extent do you feel you know the most relevant facts about energy policies?” where 1 = *not at all*, 9 = *extremely*, adapted from Rios et al., 2018). Responses were averaged to create a composite score ($M = 3.49$, $SD = 1.66$, Cronbach’s $\alpha = .92$).

Political orientation. Finally, participants responded to a demographics questionnaire assessing political orientation on a continuous measure ($M = 3.55$, $SD = 1.22$), and a categorical measure (Australian Liberal party (equivalent to Republicans in the US) = 36%, Australian Labor party (equivalent to Democrats) = 24%, Greens = 22%, other = 18%).

Results and Discussion

Differences between conditions. Similar to Study 1, a principal components analysis revealed that all three moral emotions (anger, contempt, disgust) loaded onto a single moral outrage component explaining 93% of the variance (component loadings: anger = .97, contempt = .96, disgust = .96), such that higher scores reflect greater moral outrage following one’s advocacy attempt.

Next, we proceeded to examine any differences in meta-cognitions between the two conditions. As with previous research (see Briñol et al., 2012), the meta-cognitive variables tested (perceived effort, argument quality, and source credibility) were moderately correlated ($.32 < r < .50$) and thus we performed a principal components analysis to first determine the separability of these measures. The analysis revealed that all three variables loaded onto a single component explaining 62% of the variance. This component represents general *meta-cognitive confidence* in one’s advocacy attempt (component loadings: argument quality = .46,

effort = .42, source credibility = .40), such that higher scores reflect greater meta-cognitive confidence in one's advocacy attempt. Component scores (regression method) from the principal components analysis were subjected to an ANCOVA as above.

Similar to Study 1 between-subjects ANCOVAs (Condition: Moral vs Practical) were conducted with initial stance and political orientation as covariates. Replicating the results of Study 1, moral expressiveness significantly differed between the two conditions (see Table 5-3), indicating that our manipulation is generalizable across attitude issues and sample populations. Similar to Study 1, the two conditions also differed on other moral content, and also on moral outrage. Interestingly, meta-cognitive confidence did not vary between the two advocacy conditions. This suggests that meta-cognitive confidence is unlikely to drive between-condition differences on attitude polarization. Correlations between all variables are indicated in Table 5-4 below.

Table 5-3

Means and SEs for all dictionary scores as a function of Condition (Moral versus Practical), controlling for initial stance and political orientation in Study 2

	Moral	Practical	<i>F</i>	Partial η^2
	<i>M (SE)</i>	<i>M (SE)</i>	(<i>df</i> = 214)	
Moral foundations				
Individualizing	.31(.07)	-.23(.07)	29.40**	.12
Binding	.34 (.07)	-.25(.07)	38.36**	.15
Deontology	.23(.09)	-.14(.09)	9.33*	.04
Consequentialism	-.01(.09)	.07(.09)	.40	.00
Meta-cognitive confidence	.01(.10)	-.01(.10)	.03	.00
Moral outrage	.34(.09)	-.33(.09)	28.31**	.12

Note: * $p < .05$, ** $p < .001$, two tailed tests.

Table 5-4

Pearson's Correlations Between Continuous Variables in Study 2

	POL	IS	PO	MO	MCC	CONS	DNT	BIND	IND
IND	.12	-.11	-.01	.61**	.08	.61**	.86**	.88**	
BIND	.17*	-.12	-.04	.45**	.11	.61**	.85**		
DNT	.16*	-.14*	-.09.	.34**	.05	.76**			
CONS	.16*	-.02	-.14*	-.05	.06				
MCC	.12	.06	.00	.07					
MO	.03	-.13	.16*						
PO	-.14*	-.04							
IS	.25**								
POL									

Note: **correlation is significant at the .01 level (two-tailed); *correlation is significant at the 0.05 level (two-tailed). Shorthand notation: Individualizing foundations = IND; Binding foundations = BIND; Deontology = DNT; Consequentialism = CONS; Meta-cognitive confidence = MCC; Moral outrage = MO, Political orientation = PO, Initial stance = IS; Polarization = POL.

Predictors of attitude polarization.

Polarization scores for each participant were calculated by multiplying the direction of change (time 2 attitude in the direction of time 1 attitude = 1, time 2 attitude in the opposition direction to time 1 = -1, time 1 and time 2 attitudes unchanged = 0) by the number of scale points moved from time 1 to time 2 (maximum 6 scale points). Polarization scores were created based on seminal research investigating (de)polarization in self-persuasion paradigms (see Tesser & Leone, 1977). This method is advantageous compared to pre-post difference scores, because it takes into account both the direction and extremity of attitude change, and thus provides a more complete picture of attitude change which may occur following advocacy interventions. For example, someone who initially indicated *slightly*

unfavorable towards the policy at time 1 (scale point = 3), and indicated *extremely unfavorable* towards the policy at time 2 (scale point = 1) would have a polarization score of +2 because they changed two scale points and became more extreme in their initial position (i.e. polarized). On the other hand, someone who indicated extremely unfavorable towards the policy at time 1 (scale point = 1), and indicated slightly unfavorable towards the policy at time 2 (scale point = 3), would have a polarization score of -2, because they changed by two scale points but became less extreme in their initial position (i.e. depolarized).

Interestingly, the two conditions (moral and practical) did not significantly differ on average polarization scores, $t(216) = -1.41, p = .16$. Despite the absence of a direct effect we tested the indirect effect following recommendations by previous researchers (O'Rourke & MacKinnon, 2018; Rucker, Preacher, Tormala, & Petty, 2011). In order to test our prediction that moral versus practical advocacy is consequential, we created four separate mediation models to test which of the following best accounts for variation in polarization: moral expressiveness, moral systems, moral emotions, or meta-cognitive confidence. Each model contained Condition as the independent variable, and two covariates: initial stance and political orientation.

To test the effects of moral expressiveness on polarization, we ran a mediation model using 5000 bootstrap samples and 95% confidence intervals (Hayes, 2013), with individualizing and binding foundation scores as simultaneous mediators. Results revealed a significant total indirect effect (IE) via both individualizing and binding foundations as indicated by partially standardized regression coefficients [IE = .09, SE = .04, 95% CI = .02, .16; see Figure 1 below]¹⁹. Corroborating the regression analysis, the total indirect effect was

¹⁹ Partially standardized regression coefficient is the default measurement unit in PROCESS, and controls for all variables in the mediation, but does not completely remove variation incurred by the covariates, thus taking into account natural variation in the outcome variable due to confounders (Schielzeth, 2010).

The total indirect effect of condition on polarization via moral expressiveness was significant in a mediation model without the covariates (IE = .14, SE = .06, 95% CI = .03, .27). Again, the effect appears to be driven primarily by the binding foundations (IE = .15, SE = .09, 95% CI = .00, .34).

mainly driven by the significant indirect effect via the binding foundations [IE = .14, SE = .07, 95% CI = .01, .27]^{20,21}.

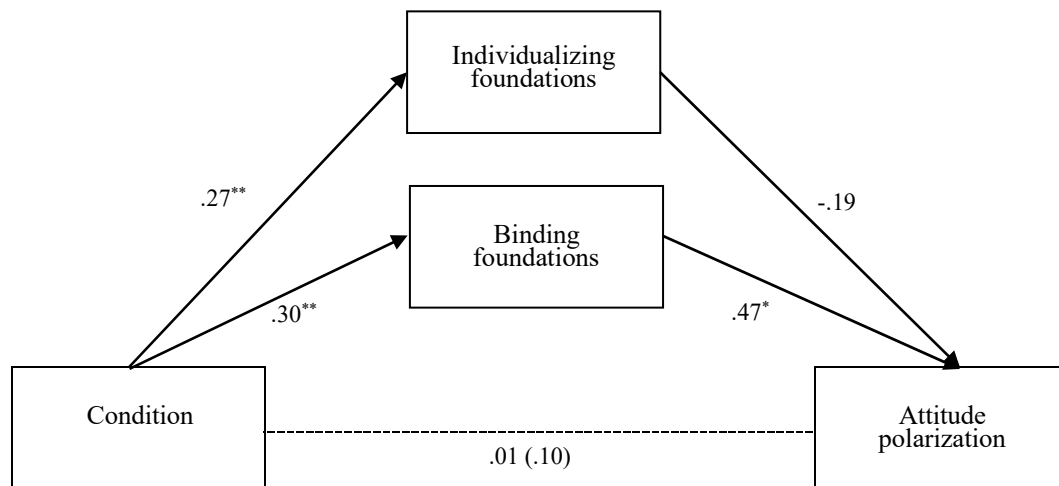


Figure 5-1. Indirect effect of Condition (moral vs practical) on attitude polarization via both indices of moral expressiveness (coding: Moral = 1, Practical = -1).

The standardized coefficient within brackets represents the total effect of Condition on attitude polarization with covariates included in the model, but without the mediators. The model controls for initial stance and political orientation but these are omitted from the diagram for clarity of presentation.

Note. * $p < .05$ ** $p < .01$.

²⁰ Post-hoc moderated mediation analyses on PROCESS (Hayes, 2013) using models 7 and 14 revealed that unlike previous work on moral framing, political orientation was not a significant moderator of the relationship between Condition and moral expressiveness or the relationship between moral expressiveness and polarization. Similarly, unlike previous work on moral matching effects, perceived moral attitude basis did not moderate the relationship between Condition and moral expressiveness or the relationship between moral expressiveness and polarization. Thus, the findings of Study 1 suggest that between-condition differences in polarization is driven by moral expressiveness, and not by other characteristics, such as political orientation, or attitude basis.

²¹ One might argue whether this effect is robust controlling for other potential mediators, such as moral outrage and deontology, which also significantly differed between advocacy conditions. Thus, we conducted a secondary simultaneous mediation analysis with the individualizing foundations, binding foundations, deontology and moral outrage as mediators, and political orientation and initial stance as covariates. Results revealed that the binding foundations continued to significantly predict polarization on its own, even after accounting for all other mediators [IE = .32, SE = .13, 95% CI = .11, .62].

Next, we tested the second proposed explanation for moral versus practical advocacy effects: moral systems framing (deontological and consequentialist reasoning). Results revealed no significant total indirect effect via both deontology and consequentialism on polarization [IE = .03, SE = .03, 95% CI = -.04, .10]. The indirect effects via deontology only [IE = .03, SE = .03, 95% CI = -.02, .10] and consequentialism only were also not significant [IE = .00, SE = .01, 95% CI = -.03, .02]. We then tested the third proposed explanation for our effects: moral outrage. Despite between-condition differences in moral outrage, results revealed no significant indirect effect via moral outrage on polarization [IE = .02, SE = .03, 95% CI = -.04, .09]. Given that meta-cognitive confidence did not vary between conditions, we did not formally test this as a mechanism in a mediation model.

Taken together, the findings of Study 2 suggest that relying on moral as opposed to practical advocacy leads to increased attitude polarization via increased moral expressiveness only, to the exclusion of other advocacy-relevant constructs, such as moral systems, moral outrage, or meta-cognitive confidence. Conversely, the findings suggest that using practical advocacy appeals attenuates attitude polarization by decreasing moral expressiveness.

In Study 3 we sought to explore other downstream consequences of moral versus practical appeals by: a) extending our dependent variables to include intentions to proselytize, and, b) testing our predictions on the original attitude issue and sample population.

Study 3

Method

Participants and design. A hundred and sixty-five participants were recruited from Amazon Mechanical Turk ($M = 35.19$, $SD = 10.62$, female = 82). This sample size was sufficient to detect the smallest effect size found in Study 1 (equivalent to $f = .30$) in One-Way ANCOVA's including two groups (Moral vs Practical) and two covariates (initial stance and political orientation) with 80% power.

Materials and procedure.

Participants followed a similar procedure to Study 1, with a few exceptions detailed below. Participants were presented with the attitude issue in Study 1. They then indicated their attitude on the same 7-point scale “(Compared to the number of migrants the government usually takes, how many migrants do you think the US government should take?”; 1 = *much fewer*, 7 = *a lot more*; $M = 4.27$, $SD = 1.79$). Those who were neutral were excluded from the study ($n = 50$) leaving a final sample of $N = 115$ (Anti = 38, Pro = 77). Participants indicated the extent to which their attitude was based on moral concerns ($M = 5.63$, $SD = 1.56$), and practical concerns ($M = 5.29$, $SD = 1.59$).

Participants were then randomly allocated to advocate for their position on the issue in practical terms ($n = 55$) or moral terms ($n = 60$) using the same task instructions used in Study 1. Participants then completed a thought-listing task prior to indicating their time 2 attitude ($M = 4.33$, $SD = 2.24$). Next, participants completed the same meta-cognitive measures as in Study 2: perceived argument quality ($M = 7.12$, $SD = 1.64$, Cronbach's $\alpha = .89$), perceived effort ($M = 7.99$, $SD = 1.11$, Cronbach's $\alpha = .82$), and perceived source credibility ($M = 5.37$, $SD = 1.85$, Cronbach's $\alpha = .93$). Participants then completed the following measures in a randomized fashion: intentions to persuade others, and perceptions of audience persuadability.

Intentions to persuade others.

Participant intentions to persuade others was assessed using five 9-point scales. Three items were taken from previous advocacy research (e.g. “How likely would you be to persuade your [family, friends, stranger] of your own position on this topic?”, taken from Cheatham & Tormala, 2015; $M = 4.96$, $SD = 2.22$, Cronbach’s $\alpha = .87$). We added two more items to measure intentions to persuade the opposition, specifically (“How likely would you be to persuade someone who disagrees with your position on this topic?”, “How likely would you be to persuade someone who holds the opposite opinion to you on the issue of migrant intake?”; $M = 4.93$, $SD = 2.41$, Cronbach’s $\alpha = .90$).

Perceptions of audience persuadability.

Participant’s perception of their advocacy appeal being effective in persuading others was assessed using a 7-point bipolar scale, “In relation to your own attitude toward the issue of migrant intake, to what extent do you think that the arguments you generated will change your target audience's attitude?” (anchored at -3 = move as far away as possible from your possible, 0 = unpersuaded/retain original position, +3 = move as close as possible to your position). This was recoded into a unipolar scale (scale points 1 to 7; $M = 4.82$, $SD = .91$).

Finally, participants completed the continuous ($M = 3.50$, $SD = 1.95$) and categorical measures of political orientation (Republican party = 25%, Democratic party = 57%, Libertarian party = 4%, other = 14%).

Results and Discussion

Differences between conditions. Corroborating the results of Study 1 and Study 2, a principal components analysis revealed that all three moral emotions (anger, contempt, disgust) loaded onto a single component explaining 92% of the variance (component loadings: anger = .95, contempt = .94, disgust = .98). Similar to Study 2, a principal components analysis revealed that all three meta-cognitive variables (argument quality,

effort, source credibility) loaded onto a single component explaining 57% of the variance (component loadings: argument quality = .92, effort = .40, source credibility = .85).

Again, replicating our results in Study 1 and Study 2, the two conditions differed on moral expressiveness, such that the moral condition generated higher scores on the individualizing and the binding moral foundations compared to the practical condition, controlling for initial stance and political orientation (see Table 5-5). Interestingly, unlike in Study 2, moral outrage did not differ between the moral and practical conditions, while unlike in Study 2, meta-cognitive confidence significantly differed between the two advocacy conditions. Corroborating the results of Study 1 and Study 2, the two conditions varied on deontology, but not on consequentialism, although this difference remains small across all studies. Correlations between all variables can be found in Table 5-6.

Table 5-5

Means and SEs for all dictionary scores as a function of Condition (Moral versus Practical), controlling for initial stance and political orientation in Study 3

	Moral	Practical	<i>F</i>	Partial η^2
	<i>M (SE)</i>	<i>M (SE)</i>	(<i>df</i> = 111)	
Moral foundations				
Individualizing	.31(.08)	-.28(.08)	27.56**	.20
Binding	.43(.09)	-.50(.09)	56.47**	.34
Deontology	.17(.13)	-.23(.12)	5.64*	.05
Consequentialism	-.12(.17)	.17(.13)	2.50	.02
Meta-cognitive confidence	.19(.13)	-.21(.13)	4.59*	.04
Moral outrage	.01(.13)	-.01(.13)	.00	.00

Note: * $p < .05$, ** $p < .001$, two tailed tests.

Table 5-6

Pearson's Correlations Between Continuous Variables in Study 3

	PRO	POL	IS	PO	MO	MCC	CONS	DNT	BIND	IND
IND	.32**	.04	.06	.15	-.00	.20*	.48**	.71**	.77**	
BIND	.31**	.01	.02	.11	-.03	.19*	.28**	.68**		
DNT	.34**	.08	-.19*		-.16	.21*	.59**			
				.32*						
				*						
CONS	.20*	.17	-.08	.17	-.12	.00				
MCC	.49**	.11	-.14	.18*	.07					
MO	.03	.08	.24*	-.15						
PO	.09	.26**	-.59**							
IS	.06	.06								
POL	.10									
PRO										

Note: **correlation is significant at the .01 level (two-tailed); *correlation is significant at the 0.05 level (two-tailed). Shorthand notation: Individualizing foundations = IND; Binding foundations = BIND; Deontology = DNT; Consequentialism = CONS; Meta-cognitive confidence = MCC; Moral outrage = MO, Political orientation = PO, Initial stance = IS; Polarization = POL; Proselytization = PRO.

Note, unlike in Study 1, although an overall regression model containing moral expressiveness, Condition, initial stance, and political orientation was significant ($R^2 = .13$, $F(5, 109) = 3.35$, $p = .007$), neither of the moral expressiveness indices predicted polarization (individualizing: $\beta = -.02$, $p = .87$; binding: $\beta = -.05$, $p = .75$). We consider explanations for this null effect in the General Discussion.

Predictors of proselytization. We now turn to our primary dependent variable in Study 3: proselytization intentions. Given: a) the conceptual similarity between intentions to persuade others, and perceptions of audience persuadability, and b) our desire to create a more parsimonious measure of intentions to proselytize, we conducted a principal components analysis to determine if these items were indeed tapping into one construct. The analysis revealed a single component explaining 74% of the variance ($.76 < \text{component loadings} < .93$). Component scores (Regression method) were used in the analyses that follow. This component collectively represents the desire to persuade others of one's position, and captures the belief that others will be persuaded (hereafter, known as proselytization intentions). The two advocacy conditions significantly differed on average proselytization scores, $t(113) = -2.15$, $p = .03$, such that those in the moral condition were more likely to express intentions to proselytize their opinions ($M = .19$, $SD = 1.06$), compared to the practical advocacy condition ($M = -.21$, $SD = .90$).

In order to test our prediction that advocacy condition indirectly influences intentions to proselytize via moral expressiveness, we conducted a mediation analysis on Process (Hayes, 2013) using 5000 bootstrap samples and 95% confidence intervals. Results revealed that the total indirect effect via both binding and individualizing foundations (combined effect) was significant (IE = .16, SE = .07, 95% CI = .03, .30; see Figure 2)²². The individual indirect effects via individualizing foundations only (IE = .07, SE = .07, 95% CI = -.06, .21)

²² The total indirect effect of condition on proselytization via moral expressiveness was significant in a mediation model without the covariates as well (IE = .16, SE = .07, 95% CI = .04, .30).

and the binding foundations only were not significant (IE = .08, SE = .09, 95% CI = -.10, .28).

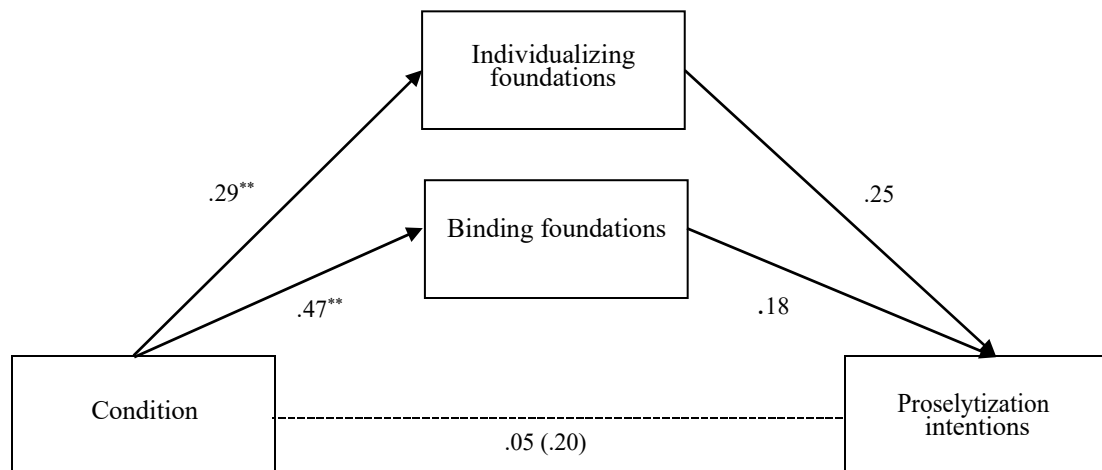


Figure 5-2. Indirect effect of Condition (moral vs practical) on proselytization intentions via both indices of moral expressiveness.

In terms of the moral systems account, results revealed no significant total indirect effect via both deontology and consequentialism on proselytization [IE = .05, SE = .04, 95% CI = -.03, .14]. The indirect effect via deontology only was marginally significant [IE = .06, SE = .03, 95% CI = .01, .13] while the indirect effect via consequentialism only was not significant [IE = -.01, SE = .02, 95% CI = -.05, .03]. We did not test moral outrage as an explanation, given the absence of between-condition differences. Finally, corroborating condition-level differences in meta-cognitive confidence, there was a marginally significant indirect mediation via meta-cognitive confidence on proselytization [IE = .09, SE = .05, 95% CI = .01, .19].

Taken together, the results of Study 3 suggest that moral advocacy may increase proselytization either via moral expressiveness, deontological framing, or meta-cognitive confidence, and not via consequentialism, or moral outrage. Across all three studies, moral

expressiveness tends to emerge as the most reliable mediator of the effects of moral advocacy on polarization and proselytization.

Mini meta-analysis A post-hoc “mini meta-analysis” was conducted across Study 2 and Study 3 following recommendations by Goh, Hall, and Rosenthal (2016), to examine the relationships between each of the two foundation clusters and polarization. Given the similarity in study design, two fixed (versus random) effect meta-analyses were conducted. The first mini meta-analysis was to assess the correlation between the individualizing foundations and polarization (Study 2 $r = .12$; Study 3 $r = .04$) and the second was to assess the correlation between binding foundations and polarization (Study 2 $r = .17$; Study 3 $r = .01$). Correlations were transformed to Fisher’s z , weighted by sample size (Study 2 $N = 218$; Study 3 $N = 115$, weight formula = $N - 3$), and then converted back to correlations for ease of interpretation.

Consistent with the results of the mediation analyses, the relationship between the binding foundations and polarization was significant, overall (Mean $r = .12$; 95% CI: .01, .22), while the relationship between the individualizing foundations and polarization was not (Mean $r = 0.09$; 95% CI: -.02, .20).

General Discussion

Across three studies, we successfully validated a novel advocacy paradigm (moral versus practical advocacy), aimed at understanding how moral language and confidence in advocacy influences post-advocacy attitudes and communicative intentions. While research in standard persuasion suggests that moral framing increases polarization and proselytization, little work has considered why this occurs. Thus, the current research examined the mechanisms through which moral rhetoric may change one’s attitudes or communicative intentions following pro-attitudinal advocacy. We tested four possible mechanisms – moral expressiveness (reliance on the five moral foundations), moral systems framing (reliance on

deontology vs consequentialist reasoning), increased emotional language use (moral outrage), and confidence in one's advocacy attempt (meta-cognitive confidence). The findings suggest that moral expressiveness is the key mechanism through which morally-framed advocacy attempts lead to polarization and proselytization. In particular, the binding foundations appears to account for the relationship between advocacy condition and polarization across both studies.

To our knowledge, this is the first study to examine the self-persuasive effects of moral versus practical messages in the context of pro-attitudinal advocacy. Our findings support previous work showing that moral rhetoric is persuasive, leads to increased proselytization, decreases willingness to compromise (Clifford, 2019; Ryan, 2017; Valenzuela et al., 2017), leads to the formation of echo chambers (Brady et al., 2017), and even leads to blatant discrimination against those holding opposing views (Pew Research Center, 2016). Our findings suggest a potential alternative to reduce polarization and proselytization: advocacy grounded in practical reasoning. We show that advocacy appeals grounded in practical argumentation leads to lower levels of moral expressiveness (as captured by both individualizing and binding moral foundations), and thus less polarization and lower intentions to proselytize.

In Study 1, we show that the practical advocacy condition generates lower levels of moral expressiveness, on average, compared to the moral condition. The practical condition also generates terms which are less semantically similar to words related to deontology, and moral outrage. In Study 2, we show that only the difference in moral expressiveness is consequential for polarization, such that, the practical condition is less likely to polarize post-advocacy, via lower levels of moral expressiveness. In Study 3, we additionally show that moral expressiveness has implications for behavioral intentions, such as, intentions to proselytize one's opinions. Importantly, those in the practical condition were less likely to

proselytize, via lower levels of moral expressiveness. Overall, the findings suggest that using practical compared moral arguments during pro-attitudinal advocacy may lead to more open communicative practices between opposing groups. Further, although previous work in standard persuasion demonstrates effects of moral values, moral systems, and emotions, in the context of self-generated persuasion (see Briñol et al., 2012; Hovland et al., 1953; Wilson, 1990), we show that the expression of moral foundations, specially, the binding foundations, is most likely to predict polarization and proselytization post-advocacy.

Moral Psychology Theory in Self-Persuasion Research

Construct distinctiveness. Although the current research treated the different moral theories (e.g. moral expressiveness and moral systems) as separate candidate explanations for our effects, the results suggest potential overlap both within and between distinct theoretical frameworks. For example, we found high correlations between the individualizing and binding moral foundations (moral expressiveness) and also between deontology and consequentialism (moral systems). In addition, we found high correlations between deontology and both moral foundations, suggesting some overlap between the different frameworks as well. This suggests that moral psychology theory may manifest differently in naturalistic expressions of moral language, compared to moral judgement tasks.

In contrast to the current work, the limited previous research using text-analysis to explore themes in moral justifications, shows no correlation between deontology and consequentialism, for example (Wheeler & Laham, 2016). This difference may be attributed to differences in linguistic processes used to justify responses to moral vignettes (which are encountered less often in daily life), compared to justifying attitudes towards controversial issues such as migration and climate-change, which is more likely to occur during conversation. Alternatively, using raw word counts as opposed to a semantic similarity

dictionary method may produce differing results, given that the latter is more sensitive to moral content.

On the hand, corroborating the high correlations observed in the current work, some previous research has shown moderate correlations between particular individualizing and binding foundations (e.g. $r = 0.40$ between care and loyalty on a climate-change issue; see Jansson & Dorrepaal, 2015). In addition, other work theorizing that harm encapsulates all the various moral foundational content (see Schein & Gray, 2015, 2018 for reviews), shows high correlations between theoretically distinct moral foundations (e.g. $r = 0.60$ between harm and loyalty; see Schein & Gray, 2015). However, the theoretical rationale for treating the individualizing and binding foundations as separate in the current research is justified given the suite of independent research evidencing the two-factor model of moral foundational content (see Davies, Sibley, & Liu, 2014; Graham et al., 2011; Nilsson & Erlandsson, 2015; Silver & Silver, 2017; van Leeuwen & Park, 2009). Whether an underlying harm foundation encapsulates all moral foundational content is beyond the scope of the current research. Nevertheless, more research is required in the context of self-persuasion to fully understand the implications of moral psychology theory in the context of self-persuasion. This is imperative, given that to our knowledge, the current research is the first to use the dictionary method to explore moral content in the context of pro-attitudinal advocacy.

Importance of the Binding Moral Foundations. Essays created by researchers in standard persuasion typically contain harm related appeals (Luttrell et al., 2019), and appeals to harm are the most common in the real-world (Clifford & Jerit, 2013). Although harm has been proposed to be the most crucial moral foundation (Schein & Gray, 2018), our studies suggest an alternative possibility. Specifically, it appears that the individualizing foundations (containing harm, care, fairness) are less important in driving polarization and in the context

of self-persuasion. Rather, advocacy condition exerted indirect effects on polarization primarily via the binding moral foundations (loyalty, authority, purity).

This suggests that targeting the binding foundations may be more effective in reducing polarization during pro-attitudinal advocacy. This notion is supported by previous work in moral reframing, which shows that presenting messages grounded in the binding foundations were more effective in changing attitudes towards recycling, than presenting messages grounded in individualizing/harm foundations (Feinberg & Willer, 2013). The authors suggest that this is because most pro-environmental messages are already framed in terms of the harm/care foundations; people already possess well-established opinions on whether something is harmful and so are unlikely to change opinions when presented with evidence in-line with harm foundations. It is possible that the binding foundations are more effective in producing depolarization because they represent less considered reasons for one's position on an issue. Indeed, previous work in self-persuasion shows that people are more likely to depolarize when presented with *novel* arguments in support of the opposition, as opposed to traditional arguments (Burnstein & Vinokur, 1977). Note, however, that the unique effect of the binding foundations did not replicate in accounting for the effect of condition on proselytization. Future research may more fully investigate effects of the moral foundations in self-persuasion contexts.

Null Effects of Political Orientation? Interestingly, we do not find any moderation by political orientation. Previous work in moral reframing (standard persuasion) consistently shows that liberals are more persuaded when presented with messages containing individualizing foundations, while conservatives are more persuaded by those containing the binding foundations (Feinberg & Willer, 2013, 2015; Voelkel & Feinberg, 2018). However, in the current study, we did not find: a) political orientation to predict average levels of moral expressiveness (liberals did not generate higher levels of individualizing foundation terms,

and vice versa for conservatives), and b) political orientation to moderate the link between moral expressiveness and polarization (liberals did not polarize more when they expressed greater levels of individualizing foundation terms, and vice versa for conservatives).

Although previous work in standard persuasion suggests “fundamental moral differences separating liberals and conservatives” (Voelkel & Feinberg, 2018) we do not find these differences in the context of self-generated persuasion, suggesting differences in standard versus self-persuasive processes. Indeed, recent work has shown that liberals and conservatives rely on similar moral foundations in making moral judgements of influential figures, suggesting that differences between political partisans may be exaggerated in moral foundational theory (Frimer, Biesanz, Walker, & MacKinlay, 2013). It is possible that political orientation plays a greater role during standard persuasion, but its effects are attenuated in the context of self-generated persuasion, whereby the content of the message generated is more important than demographic characteristics.

However, it is important to note that political orientation significantly correlated with particular moral content in the current research, although these effects were not consistent between and studies, nor large enough to moderate the effects of content on polarization or proselytization. For example, political orientation significantly negatively correlated with both moral foundations and deontology in Study 1, suggesting a link between conservatism and a lower tendency to express content related to these two moral domains. To consider another example, political orientation significantly positively correlated with moral outrage in Study 2, suggesting a link between conservatism and expression of moral outrage. These findings suggest that political orientation may play a more subtle role in the context of self-persuasion, compared to standard persuasion, and presents an interesting direction of research which is yet to be explored in the context of advocacy.

Potential differences between standard and self-persuasive processes is further highlighted by the lack of “moral matching” effects in the current study. Previous work shows that presenting counter-attitudinal moral messages (compared to practical messages) tend to be more effective in persuading those who have a moral attitude basis (see Luttrell et al., 2019). However, this effect does not occur in the current research, again suggesting that demographic characteristics and attitude properties may be less important in the context of self-generated persuasion, compared to characteristics of the message itself, such as, moral expressiveness. Future research is required to affirm this claim.

In sum, we show that practical reframing of advocacy appeals may depolarize attitudes and decrease proselytization via self-persuasion. Our findings suggest that moral reframing and moral matching effects may be less effective in reducing close mindedness in the context of pro-attitudinal advocacy. Importantly, the difficulties encountered with encouraging collaboration between liberals and conservatives (Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014) may be combated by encouraging advocacy grounded in practical as opposed to moral reasoning, regardless of one’s initial stance or political orientation. These findings carry implications for re-designing online discussion forums (e.g. Reddit), and informing policy recommendations around important socio-political issues such as migration and climate-change.

Limitations and Future Directions

Failure to Replicate Polarization-Moral Expressiveness Link

A limitation of the current research is that the indirect effect of advocacy type (moral vs practical) on attitude polarization via moral expressiveness failed to replicate in Study 3. One possibility is that the effect of moral expressiveness on polarization was detected in Study 2 but not in Study 3 due to a smaller sample size. The post-hoc mini meta-analysis revealed that the relationship between the binding foundations and polarization was

significant overall, while the relationship between the individualizing foundations and polarization was not. Although the relationship between moral advocacy and polarization can be explained by the expression of binding moral values, this is less clear in explaining the effects of advocacy on proselytization, as neither foundation type predicted proselytization on their own. Future work is required to disentangle the effects of advocacy on the various dependent variables using larger samples from diverse populations.

Another explanation for the diluted effect of polarization in Study 3, is differences in the type of attitude issue tested, as issues vary in their controversy and propensity to evoke emotional reactions (Reeves et al., 2016). It is likely that the issue tested in Study 3 (migration) is more affectively-laden and more morally convicted compared to the issue tested in Study 2 (carbon emissions policy, relatively unfamiliar issue). Indeed, a post-hoc *t*-test shows that attitudes towards migration in Study 3 were on average, significantly more based on moral concerns, compared to attitudes towards the carbon emissions policy ($p < .001$, Cohen's $d = 0.78$ medium effect size). This suggests that attitudes towards migration may be more morally ingrained compared to issues related to climate-change (at least in the sample tested), making them more difficult to shift (Krosnick & Petty, 1995).

This is congruent with early research in self-persuasion, showing that depolarization of attitudes tends to be difficult to achieve, in general, (Tesser, 1978), and even more so for attitudes which are highly morally convicted (e.g. capital punishment; Vinokur & Burnstein, 1978). Nevertheless, even on a highly convicted issue, we were able to observe shifts in proselytization following our manipulation, which is quite promising. Encouraging the use of practical reasoning compared to moral reasoning resulted in lower intentions to proselytize, via lower levels of moral expressiveness, even if attitudes did not change. Future research may examine other downstream consequences of practical versus moral advocacy on communicative intentions such as, willingness to interact with those holding opposing views.

Lack of “Control” Condition

It may be argued that the lack of a true control condition limits the inferences which can be drawn about the self-persuasive effects of advocacy on the outcome variables (depolarization and proselytization). The primary aim of this research was to consider differences specifically between moral and practical advocacy, as opposed to deviation from a control (for similar study designs in pro-attitudinal advocacy which employ two or more experimental conditions without a control condition, see Briñol et al., 2012; Gordijn et al., 2001). Nevertheless, given the lack of a control condition, our inferences are limited to the effects of moral compared to practical advocacy only. Future research may replicate this research with a control condition, in order to more fully draw out the implications of different types of pro-attitudinal advocacy. Further, future research may also investigate whether the effects of moral versus practical advocacy are limited to self-persuasion only, or whether they may apply to standard persuasion as well (i.e. when arguments are presented to participants rather than written by them). We believe that this is a useful direction of research opened up by the current work.

Inability to Draw Causal Inferences

We note that the correlations observed in the current work do not provide strong evidence for causal mediation. This is an inherent limitation of using cross-sectional mediation analysis as a test of causal process (see Kline, 2015). Although we find that advocacy indirectly predicts polarization and proselytization via moral expressiveness, future research is required to demonstrate causal links between these variables. Similar to the current work, even though past research typically manipulates an advocacy context, it tends to investigate the relationships between dependent variables in a correlational manner (see Briñol et al., 2012 for work on pro-attitudinal advocacy and meta-cognitive confidence). Future work manipulating the mediator (e.g. manipulating individualizing and binding

foundation levels) is required to establish moral expressiveness as the key driver of our effects. Nevertheless, this research highlights important initial evidence that moral expressiveness is associated with polarization and proselytization, specifically in the context of pro-attitudinal advocacy.

Conclusions

We find evidence that moral advocacy may lead to increased social division, via greater expression of moral foundations content, compared to other types of moral content, such as deontology, or moral emotions. We provide evidence for a novel pro-attitudinal advocacy manipulation, i.e. practical advocacy, which may promote depolarization and reduce proselytization via self-persuasion. Our findings suggest that encouraging advocates to generate practical reasons for their attitude leads to less polarization and lower intentions to proselytize one's attitude. The findings provide evidence for a novel and effective method of encouraging free speech, while simultaneously minimizing social conflict.

It is important to note that moral discourse, especially about charged political topics, may have its own value in promoting social cohesion and enabling corrective action, depending on the context. For example, recent research finds that moral language is more persuasive than economic language (i.e. monetary arguments) to influence company management on important social issues faced by employees (Mayer, Ong, Sonenshein, & Ashford, 2019). That is, presenting arguments framed in moral terms as opposed to economic terms is more effective in mobilizing decision-makers on social issues such as, employee health, gender equality, and wage policies. Thus, while practical self-persuasion may be a path to greater harmony, the inverse may also be true: moral arguments may mobilize society in ways that could be practically important. That is, there may be contexts in which each outcome (harmony vs. mobilization) is desirable. Future research may more fully examine

the potential of both moral and practical advocacy to bridge ideological divides, by enabling open communication between fragmented social groups.

Funding

The authors have no funding to report for research design, execution, analysis, interpretation and reporting. An Australian Government Research Training Program Scholarship was bestowed to the corresponding author for the completion of their doctoral dissertation.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability

The experiments conducted in the research were not pre-registered. The data files used for analysis have been made available on the OSF repository.

<https://osf.io/nd6vs/>

Acknowledgements

We thank Manuel José Rengifo and Shaheed Azaad from the Melbourne School of Psychological Sciences, for their useful suggestions regarding the mini-meta analysis.

Supplementary Materials

Appendix A

Attitude issue descriptions.

Migration. “People migrate for a variety of reasons, including, to study, to work, to reunite with family, or to escape violence and persecution in their home country. There is currently a migration crisis worldwide, however, host countries have limitations with respect to how many migrants they can accommodate. Decisions with respect to migrant intake are detailed in migration policy. Recently, governments and political parties have expressed different views on migrant intake quotas”.

Carbon emissions.

“Carbon dioxide emissions from electricity and gas usage contribute to more than half of Victoria’s total carbon dioxide emissions. These emissions peak particularly during colder times of the year, due to significantly increased usage of heating. The higher the household thermostat temperature, the greater the energy consumed. The Australian government may therefore pass a law specifying that households set a maximum thermostat temperature of 20 degrees Celsius during winter.”

Appendix B

Advocacy task instructions.

Advocacy task (Moral).

While our focus was tapping into the five moral foundations (harm, fairness, loyalty, authority, purity) (see Haidt, 2007; Haidt, 2001), instructions for the moral condition were created based on both moral psychology and positive psychology, to capture the broad range of values and virtues which tap into morality (see Peterson & Seligman, 2004 for moral virtues). This was in order to ensure that participants were not constrained in the expression

of their attitudes. In the moral advocacy task, participants were encouraged to think about a moral choice as "weighing issues of harm, rights, justice and fairness", considering rights violations, and making decisions "with respect to a set of virtues in a culture or subculture" (Haidt, 2001). To capture these concerns, our index of moral expressiveness was scored on the binding moral foundations (average scores on loyalty, authority, purity) and individualizing moral foundations (average scores on harm and fairness; see Jonathan Haidt & Joseph, 2004, 2008). Participants were presented with the following instructions:

"People use different kinds of arguments to persuade others. For example, they may use arguments grounded in moral values. On the other hand, they may use arguments grounded in practical concerns, such as, resource availability. For this task, we ask you to focus on fundamental moral values which relate to your position on migration. We are not interested in practical concerns related to your position. Imagine that you have to persuade someone who disagrees with you on your view about how many migrants the American government should take. Convince this person to adopt your own position on this issue. Specifically, write arguments **only** from a **moral** perspective. That is, base your arguments on moral considerations such as (but not restricted to) the following:

- Whether something is fundamentally right or wrong, good or bad.
- Whether something is worthy of blame or punishment or moral praise.
- Moral values relating to concerns such as care, harm, injustice, fairness, loyalty, betrayal, disrespect, purity or liberty.
- Virtues such as wisdom, courage, generosity, transcendence, justice, prudence".

Advocacy task (Practical). Instructions for the practical condition were created based on economic theory, to capture concerns related to resource scarcity. In the practical advocacy task, participants are encouraged to think about a practical or rational choice as one that "maximizes expected utility" (Kahneman & Tversky, 1986; Tversky & Kahneman,

1981). It is thinking in terms of outcomes (e.g. if X happens then Z consequences), in terms of money, and resources, for example. Participants were told that a practical choice is on which allocates scarce resources to the alternative which maximizes positive outcomes for the group; “the rational man is a maximiser” (Simon, 1978). It is balancing consequences (i.e. cost-benefit analysis) to pursue goals which “maintain the integration and adaptation of the larger system” (Simon, 1978).

“People use different kinds of arguments to persuade others. For example, they may use arguments grounded in moral values. On the other hand, they may use arguments grounded in practical consequences, such as, resource availability.

For this task, we ask you to focus on practical concerns which relate to your position on migration. We are not interested in moral values related to your position. Imagine that you have to persuade someone who disagrees with you on your view about how many migrants the American government should take. Convince this person to adopt your own position on this issue.

Specifically, write arguments **only** from a **practical** perspective. That is, base your arguments on practical considerations such as (but not restricted to) the following:

- Whether something is feasible or workable, pragmatic or realistic
- Whether something is based on evidence, on something concrete and factual
- Whether something is grounded in rational cost-benefit analyses (e.g. costs outweighing benefits, or benefits outweighing costs)
- Consider things like resource considerations (e.g., how resources are allocated), economic consequences.”

Appendix C

Constructing dictionaries.

In constructing the dictionaries, we generally aimed to avoid words with common but irrelevant alternate meanings. For example, the word right (in the sense of “human right”) is relevant to deontological ethics, however the word is arguably more commonly used to reference either correctness (e.g., “the right answer”) or location (e.g., “to the right”). Because some dictionary categories were represented by a very small number of terms, we performed a wildcard search through the model vocabulary to identify additional variants of our keywords. This strategy serves the goal of boosting the reliability of the dictionary by including vectors for relevant variants of a word (much like adding additional items to a questionnaire). Additionally, by including variants of the same word we are able (to some extent) to average out idiosyncratic or unwanted alternate meanings of some of the words. For example, in the case of the dictionary intended to capture expressions of Contempt, by including multiple variants of the word contempt (e.g., contemptible), the common use of contempt as a legal term (as in “contempt of court,” which doesn't reference any actual emotional state) should have relatively less influence on the concept representation.

To identify the words that would define each concept, we went through the original W&L dictionary, and identified the most important words for each dictionary category. For each of these words, we performed a wildcard search through the model vocabulary so we could average over relevant variants in the dictionary (e.g., averaging over variants of *anger*, *rage*, etc.) before averaging over the composite vectors to derive our representation of Anger. The rationale for the two-step averaging (within word stems and then across word stems) rather than just averaging all terms together was that different words had vastly different numbers of variants (e.g., there were two relevant variants of furious, but five relevant variants of anger), however we didn't want the category vectors to be

substantially weighted in favor of whatever words happen to have the most variants, because it's unlikely that the number of variants of a word corresponds to their importance for our dictionary categories. After generating the wildcard list, it was manually inspected by the authors who decided which words to include in the final dictionary.

For the emotion dictionary concepts, we decided to omit antonyms, because although such terms are obviously semantically related, their inclusion would be inconsistent with the goal of creating dictionaries that capture the expression of the particular emotion. For the non-emotion dictionaries however, we decided to include antonyms (e.g., having responsibility and irresponsibility both representing Deontology) because both arguably express consideration of that particular moral code.

The full list of words used to capture all dictionary concepts are indicated in the table below:

Concept	Words
Moral foundations	
Harm	suffer
	cruel
	hurt
	harm
Care	kindness
	compassion
	nurture
	empathy
Fairness	fairness
	equality
	justice

	rights
Cheating	cheat
	fraud
	unfair
	injustice
Subversion	subversion
	disobey
	disrespect
	chaos
Loyalty	loyal
	solidarity
	patriot
	fidelity
Betrayal	betray
	treason
	disloyal
	traitor
Authority	authority
	obey
	respect
	tradition
Purity	purity
	sanctity
	sacred
	wholesome
Degradation	impurity
	depravity

	degradation
	unnatural
Moral systems	
Deontology	duty
	obligation
	obligations
	obligated
	obligatory
	obligate
	obligates
	obligating
	rights
	righteous
	righteousness
	rightful
	rightfully
	righteously
	human-rights
	civil-rights
	rights-based
	responsible
	responsibility
	responsibilities
	irresponsible
	responsibly
	irresponsibility
	irresponsibly

	responsibility
	responsibility
	responsibilities
	responsibilities
	rules
	rule
	rule-breaking
	norm
	norms
	normative
Consequentialism	outcome
	outcomes
	outcome-based
	consequences
	consequence
	consequential
	consequent
	consequentially
	consequentialism
	consequentialist
	repercussion
	repercussions
	repercussion
	cost
	costs
	costly
	cost-effective

costing

low-cost

cost-effectiveness

cost-effectively

cost-benefit

cost-efficient

costlier

lower-cost

high-cost

cost/benefit

costliest

cost-based

cost-efficiency

lowest-cost

benefits

benefit

beneficial

benefited

beneficiaries

beneficiary

benefiting

benefitted

benefitting

beneficent

beneficially

beneficiating

beneficence

	risk-benefit
	benefic
	benefit-cost
Moral emotions	
Anger	angry
	anger
	angered
	angers
	angering
	outrageous
	outrage
	outraged
	outrageously
	outrages
	outrageousness
	furious
	furiously
	rage
	enraged
	rages
	raged
	enrage
	frustrated
	frustrating
	frustration
	frustrations
	frustrate

	frustrates
	frustratingly
Contempt	contempt
	contemptuous
	contemptible
	contemptuously
	despise
	despised
	despises
	despising
	detest
	detested
	detestable
	detests
	detestation
Disgust	disgusting
	disgust
	disgusted
	disgustingly
	disgusts
	repulsive
	repulsed
	repulsion
	repulse
	revolting
	sickening
	sickened

sickens

sicken

sickeningly

Chapter 6 Attack vs Defend: Advocacy Style and Self-Persuasion

Preface to Chapter 6

(This chapter was previously submitted to *Basic and Applied Social Psychology* as “Effects of Attack versus Defense Advocacy Styles on Uncongeniality Bias and Openness to Opposing Views”; the submitted version has been included in this thesis)

Similar to Chapter 5, we here contrasted two distinct types of pro-attitudinal advocacy which advocates may engage in: Pro-attitudinal Defense (defending one’s position) versus Counter-attitudinal Attack (attacking the opponent’s position). We propose these two types of pro-attitudinal advocacy by drawing inspiration from interdisciplinary research. For example, previous work in political advertising shows that political candidates typically engage in negative advertising (i.e. degrading the opposition) compared to positive advertising (i.e. highlighting their positives) to gain support from voters (see Moorman & Neijens, 2012 for a review). In addition, the communications literature distinguishes between *advocating* (substantiating one’s position) and *refuting* (opposing an alternative position), while demonstrating different propensities for individuals to engage in one or the other (see Infante & Rancer, 1993; Kuhn, 1991).

However, neither political advertising literature nor the communications literature empirically demonstrate the effects which advocating versus refuting behaviors have on the *advocator’s* own subsequent attitudes and behaviors. Thus, the current research extends this work to pro-attitudinal advocacy, given the preliminary differences demonstrated between advocating and refuting and the limited experimentally manipulated pro-attitudinal advocacy conditions in social psychology (see Briñol et al., 2012; Gordijn et al., 2001 for exceptions). Thus, we make a novel contribution to pro-attitudinal advocacy research, by proposing Defense versus Attack advocacy styles as distinct styles of advocacy.

Given the limited investigation of pro-attitudinal advocacy, we were unable to make concrete predictions about: a) the direction of effects of Defense and Attack advocacy styles on a novel set of communicative intentions, such as information-seeking preferences, and b) the specific mechanism underlying any effects. Thus, we proposed three competing alternative hypotheses for the direction and mechanism of our effects: (1) meta-cognitive confidence-driven self-persuasion, (2) emotions, and (3) accuracy versus defense motivations (see Hart et al., 2009 for a review). Interestingly, our research reveals evidence for a mechanism other than for self-persuasion in accounting for the effects of Defense versus Attack on information-seeking preferences and willingness to engage with dissimilar others.

Citation: **Abeywickrama, R. S., & Laham, S. M. (2020b).** Effects of Attack versus Defense Advocacy Styles on Uncongeniality Bias and Openness to Opposing Views. *Basic and Applied Social Psychology* (in progress).

**Effects of Attack versus Defense Advocacy
Styles on Uncongeniality Bias and Openness to
Opposing Views**

Abstract

The current research proposes two distinct styles of pro-attitudinal advocacy which may influence communicative intentions relevant to bridging gaps between opposing groups: *Pro-attitudinal Defense* (defending one's own position) versus *Counter-attitudinal Attack* (attacking an opponent's position). Across two studies ($N_1 = 125$, $N_2 = 178$) those who advocated for their position by attacking opponents' views were found to be more likely to display open communicative intentions, such as, selecting information and interaction partners with opposing views to their own when compared with Defense advocates. This research suggests that deliberating about the opposition's position may induce more open communicative practices in one's self than elaborating on one's own position might, by inducing a motivation to hold more accurate worldviews. We explored multiple explanations, and end by discussing the implications of harnessing pro-attitudinal advocacy to bridge attitudinally-polarized groups.

Keywords: attitudes, persuasion, self-persuasion, advocacy, depolarization

Introduction

Over the past decade, the United States has become increasingly polarized in its attitudes towards socio-political and moral issues (Pew Research Center, 2016). Despite reporting receptivity to attitudinal diversity, people continue to express prejudice towards, and distance themselves from, those holding different worldviews (Brandt, 2017; Hart et al., 2009). Attitude polarization tends to be exacerbated by communicative practices; for instance, discussing one's attitude in a group typically leads to more extreme individual attitudes (group polarization; see Myers & Lamm, 1976; Sunstein, 2002 for reviews). More generally, advocating one's own position (i.e. pro-attitudinal advocacy), even in the absence of a group, leads to attitude polarization as well (see Briñol et al., 2012; Gordijn et al., 2001). That is, individuals can become polarized even by their own advocacy behaviors.

Despite its prevalence and importance in constructing the fractured attitude landscape (especially in online settings; Hestres, 2015), little work has systematically examined pro-attitudinal advocacy. In the current work, we consider the role of two distinct types of pro-attitudinal advocacy, which we term Counter-attitudinal Attack (attacking an opponent's position; hereafter *Attack*), and Pro-attitudinal Defense (defending one's own position; hereafter *Defense*) in shaping attitudes and communicative intentions. Although no work - to our knowledge - has directly examined the effects of Attack versus Defense advocacy styles on attitude polarization and communicative intentions in the context of pro-attitudinal advocacy, this distinction mirrors one in political advertising between positive (highlighting one's own merits) and negative (showcasing opponents' flaws) advertising techniques (see Moorman & Neijens, 2012 for a review). Additionally, the distinction between Attack and Defense advocacy styles has been captured in the communications literature – people tend to

argue with others by advocating (elaborating one's own position) or refuting (attacking the opponent's position; see Infante & Rancer, 1993).

In the current research, we explore the effects of attacking a position held by others versus defending one's own position on attitudes and behaviors relevant to bridging gaps between polarized groups, such as, seeking attitude-inconsistent information and expressing willingness to select interaction partners with different views.

Self-Persuasive Effects of Pro-Attitudinal Advocacy

In spite of its important place in understanding the dynamics of echo chambers and political polarization (Leifeld, 2014), little work has been undertaken in advancing knowledge on the self-persuasive effects of pro-attitudinal advocacy. Thus, we compare and contrast the attitudinal and behavioral consequences of two different types of pro-attitudinal advocacy styles: Attack and Defense. Following a pro-attitudinal advocacy attempt, an advocator may unconsciously "self-persuade" themselves to hold a more or less extreme attitudinal position post-advocacy (self-generated persuasion; Briñol et al., 2012). That is, the advocator may polarize or depolarize post-advocacy. Furthermore, in addition to changing one's attitudes via self-persuasion, one's behavioral intentions may also change.

Interestingly, changes in behavioral intentions may occur even under circumstances in which attitude change does not take place, especially when attitudes are controversial (Krosnick & Petty, 1995) and held with high conviction, a phenomenon prominently evidenced in activity concerning attitudes towards migration or climate-change (Abeywickrama, Laham, & Crone, 2018; Hobson & Niemeyer, 2013; McDonald & Ryan, 2019).

The scant existing literature on this subject points to a few conditions under which attitude and behavior change may occur as a result of pro-attitudinal advocacy. For instance, attitude depolarization has been shown to occur more frequently when advocates generate arguments for a minority opinion as opposed to a majority opinion (Gordijn et al., 2001).

Attitude polarization has also been shown to occur when advocates generate arguments grounded in moral compared to practical reasoning, depending on specific mechanisms considered such as, moral values, meta-cognitions, and emotions (Abeywickrama et al., 2020). Significantly, recent work in pro-attitudinal advocacy demonstrates that while advocacy may not influence attitude polarization itself, it may influence both “open” communicative intentions, such as intentions to engage with those holding dissimilar views, and “closed” communicative intentions, such as willingness to proselytize one’s existing held opinions (Abeywickrama et al., 2020; Abeywickrama & Laham, 2020a).

Other than for the research surrounding these effects, little is known about the kinds of variations in pro-attitudinal advocacy which lead to attitude depolarization and subsequent open communicative intentions. Thus, we propose another dimension along which pro-attitudinal advocacy can vary – Attack versus Defense.

Attack vs Defense Advocacy Styles

The dimension proposed has been conceptualized outside social psychology in both political advertising and communications literatures. In political advertising research, for instance, presidential nominees are described as using “attack politics” to garner support by pointing out competing candidate’s flaws (Haynes & Rhine, 1998). A similar distinction is drawn in the communications literature as a paradigm of “advocating” versus “refuting”, in which the former involves elaborating one’s own position while the latter involves attacking an opponent’s position (see Infante & Rancer, 1993; Kuhn, 1991).

However, the little existing work which has examined this distinction tends to focus exclusively on the antecedents of employing these advocacy styles. As an example, when given the opportunity to argue with a confederate on a controversial topic, adults are more likely to refute (Attack) as opposed to advocate (Defend) (Limon, Turner, & Zompetti, 2008), while the converse is true for teenagers (Felton & Kuhn, 2001; Kuhn, 1991). Likewise, those

high in trait-level argumentativeness are more likely to use both Attack and Defense arguments to substantiate their position compared to those who are low or moderate in argumentativeness (Infante & Rancer, 1993).

Attack and Defense has been conceptualized previously in social psychology itself, specifically in persuasion research investigating the types of strategies people may use to resist persuasive attempts from external sources (Jacks & Cameron, 2003; McGuire, 1964). This work has identified and directly contrasted two common strategies used to resist persuasion: *counter-arguing* (rebuttal of others' arguments; akin to Attack), and *attitude bolstering* (generating arguments consistent with one's attitude; akin to Defense). This work demonstrates that compared to other persuasive strategies (e.g. derogating the message source, drawing on group consensus), counter-arguing and attitude bolstering are used more commonly in argumentation (Jacks & Cameron, 2003). This work crucially demonstrates variations in the impact of advocacy style on subsequent attitudes, such that only counter-arguing has been shown to predict attitude resistance (minimal attitude change pre vs post-advocacy) in response to a counter-attitudinal message (Jacks & Cameron, 2003).

While illuminating, the little existing work distinguishing these two advocacy styles is limited in that it: a) focuses mainly on antecedents of advocacy (e.g. effects of trait-level argumentativeness; see Rancer & Avtgis, 2006 for a review), or frequency of style of advocacy (Felton & Kuhn, 2001; Jacks & Cameron, 2003; Kuhn, 1991; Limon et al., 2008), b) is limited in explicating the consequences of advocacy on people's subsequent attitudes and behavior (see Jacks & Cameron, 2003 for an exception), and c) tends to examine natural variations in advocacy, as opposed to experimentally manipulating advocacy styles, thus limiting causal inference. Thus, in the present research we manipulate Attack and Defense advocacy styles to examine the downstream consequences on attitude change and communicative intentions.

Advocacy Effects on Attitudes and Communicative Intentions

As noted previously, while pro-attitudinal advocacy may influence attitude change directly (Abeywickrama & Laham, 2020a; Briñol et al., 2012; Gordijn et al., 2001), it may also simply influence communicative intentions, such as, willingness to engage with dissimilar others and proselytize one's opinions, even in instances where attitude change itself does not necessarily occur (Abeywickrama et al., 2020; Abeywickrama & Laham, 2020a). Thus, in the current work, we place greater emphasis on communicative intentions, given the limited work examining the links between pro-attitudinal advocacy and behavior.

For the purpose of this research, we identify and construe communicative intentions and practices on a continuum ranging from 'open' to 'closed.' Open communicative practices include those which promote greater engagement between polarized groups; such as uncongeniality bias, as indexed by seeking attitude-inconsistent information (see Hart et al., 2009 for a review) or choosing interaction partners with opposing viewpoints (Rios et al., 2018). Open communicative practices can also be captured by assessing the extent to which advocates express increased receptivity to opposing views (Abeywickrama & Laham, 2020a; Voelkel & Feinberg, 2018).

In contrast, closed communicative practices include those which may exacerbate distance between opposing groups, such as, seeking attitude-consistent information (congeniality bias), expressing intentions to proselytize one's opinions (persuasion intentions; Abeywickrama et al., 2020; Cheatham & Tormala, 2015, 2017) or choosing interaction partners with similar opinions to one's own.

We contend that communicative intentions and practices may become more open or closed depending on whether people employ Attack or Defense advocacy styles to advocate for their opinions. Specifically, we predict that advocacy style may influence *uncongeniality*

bias (as indexed by advocates selecting attitude-inconsistent information or choosing interaction partners with opposing views) and also a more general *open mindset* (as indexed by expressing receptivity to opposing views and perceiving the opposition as trustworthy and competent).

Although there are many theoretical perspectives we could take to examine the mechanisms underlying the relationship between advocacy style, attitude polarization, communicative intentions, and open mindsets, we are particularly interested in the effects of meta-cognitions and emotion. The reasoning for our choice of mechanisms is three-fold, as outlined here:

First, limited existing work in social psychology points to meta-cognitions as a potential mechanism for the downstream consequences of advocacy. For instance, perceived knowledge has been shown to predict the use of particular advocacy styles, such that higher perceived knowledge predicts greater likelihood of using counter-arguing (i.e. Attack) due to greater confidence in one's arguments (Jacks & Cameron, 2003). More importantly, meta-cognitions have been shown to not only act as antecedents of advocacy styles, but to explain the links between advocacy styles and attitude change. For instance, only counter-arguing (i.e. Attack) and not attitude bolstering (i.e. Defense) has been shown to predict attitude resistance, and this effect is qualified by thought favorability (whether the advocate generated unfavorable thoughts that directly refuted the counter-attitudinal position; Jacks & Cameron, 2003).

Second, consistent with these findings, recent research in pro-attitudinal advocacy has demonstrated differences between advocacy styles specifically on the grounds of meta-cognitions and emotions. As an example of this, pro-attitudinal advocacy grounded in moral arguments has been shown to generate higher levels of meta-cognitive confidence and approach-motivated emotions such as, anger, compared to advocacy grounded in practical

arguments (Abeywickrama, Rhee, Crone, & Laham, 2020). Importantly, these effects have been shown to predict communicative intentions, as in cases where meta-cognitive confidence in one's advocacy attempt has been shown to directly predict open communicative intentions, such as, willingness to engage with those holding dissimilar views (Abeywickrama & Laham, 2020a).

Finally, our focus on specific constructs such as meta-cognitions and emotions endeavors to bridge the gap between political advertising, communications research, and social psychology, while applying a novel self-persuasion and pro-attitudinal advocacy lens to the traditional literature conceptualizing Attack (i.e. "negative advertising", "refuting", "counterarguing") and Defense ("positive advertising", "advocating", "attitude bolstering") in its various forms. The novelty of this research is highlighted by our experimental manipulation of advocacy styles, while exploring its effects on both attitude change and communicative intentions.

Hypotheses

We predict that Attack and Defense advocacy styles may influence attitudes and communicative intentions, in particular, uncongeniality bias and open mindsets. While we put forward mechanisms grounded in meta-cognitions and emotions, given the limited work in Attack versus Defense, we are unable to make specific predictions regarding the direction of effects. However, we are able to generate some hypotheses based on extant research in self-persuasion. We explore these hypotheses in-turn.

Meta-Cognition Hypothesis

One possibility is that advocacy style influences one's attitudes and communicative intentions by changing meta-cognitive confidence (i.e. how confident advocates are in their persuasive attempt; see Petty et al., 2007 for a review). Extant work in self-persuasion demonstrates that people find it difficult to justify their own attitudes (Wilson & Dunn,

1986), suggesting that Defense may be more difficult than Attack. Difficulty justifying one's attitudes may lead to Defense advocates displaying lower meta-cognitive confidence, as indexed by various confidence-related variables, such as, perceived quality of one's arguments (Briñol et al., 2012), perceived effort expended in generating those arguments (Briñol et al., 2012), favorability of thoughts generated during advocacy (thought favorability; ratio of positive to negative thoughts generated during advocacy; see Clark, Wegener, Sawicki, Petty, & Briñol, 2013), and confidence placed in thoughts generated during advocacy (thought confidence; see Petty, Briñol, & Tormala, 2002). This lower meta-cognitive confidence displayed by Defense advocates may then lead to more open communicative intentions, with or without influencing attitude depolarization itself (Abeywickrama & Laham, 2020a). Conversely, given that advocates: a) often have difficulty generating reasons to support their own attitudes (Wilson & Dunn, 1986), and b) tend to rely on refuting opponents rather than justifying their own position (Limon et al., 2008), a meta-cognition account may predict that Attack advocates will display more closed communicative intentions, via heightened meta-cognitive confidence and/or attitude polarization.

On the other hand, meta-cognition may also elicit the opposite effects of Attack on attitude polarization and communicative intentions. In contrast to work in communications research, work in persuasion resistance suggests that counterarguing advocacy strategies are perceived to be used less commonly than attitude bolstering strategies (Jacks & Cameron, 2003). If this is true, in the context of pro-attitudinal advocacy, then Attack advocates may have difficulty generating arguments due to relatively lesser previous experience advocating compared to Defense counterparts, resulting in lower confidence in advocacy attempts of the former group. This lower meta-cognitive confidence may then predict greater uncongeniality bias, such as, seeking attitude-inconsistent information or displaying greater receptivity to opposing views, given understanding of previous work in pro-attitudinal advocacy

demonstrating greater openness as a function of lower meta-cognitive confidence (Abeywickrama & Laham, 2020a).

In sum, we may expect the effects of advocacy style on communicative intentions (regardless of direction) to be driven by meta-cognition and/or attitude polarization.

Emotion Hypothesis

In contrast to some self-persuasion and communications research, emotions research suggests, somewhat paradoxically, that Attack may lead to more open communicative intentions via increased ‘confrontational’ emotions. Two hypotheses follow from this. First, compared to Defense, Attack may be more likely to elicit confrontational emotions, such as, anger, as attacking the opposition may heighten the incongruence between the attitudes of the advocate and opposition, which may be perceived as a threat to each party’s held values (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013), and thus productive of anger. Second, confrontational emotions, including anger, have been shown to be associated with more open communicative practices, such as seeking attitude-inconsistent information.

As an illustration, in an emotion induction task, angry participants actively chose more attitude-inconsistent information compared to sad participants, thus effectively overcoming congeniality bias (Young, Tiedens, Jung, & Tsai, 2011). In addition, angry participants have been shown to display superior perspective-taking skills in advocacy tasks, as indexed by greater generation of attitude-inconsistent arguments compared to neutral participants (Kim, 2016). This effect might be explained by the induction of anger of an increased willingness to scrutinize others’ ideas (Wiltermuth & Tiedens, 2011). What this means is that Attack advocates may exhibit greater uncongeniality bias than those in the Defense condition, as such information presents them with an opportunity to garner ammunition against the opposition (Young et al., 2011). Thus, emotions research suggests

that Attack may lead to more open communicative intentions than Defense, and that this effect will be accounted for by anger and/or heightened emotional arousal.

In contrast to work in the emotions literature, work in communications research suggests that anger may elicit the opposite effects of Attack on communicative intentions. Specifically, this work implies a case for psychological reactance (i.e. how people respond when their freedom and ideologies have been threatened; Brehm & Brehm, 1981). It is possible that Attack induces greater psychological reactance (increased threat to one's freedoms and attitudes) by highlighting the incongruence between the opposition's position and one's own. This may lead to feelings of anger and/or increased emotional arousal, given that psychological reactance is typically characterized by anger and the presence of counter-arguing thoughts (Dillard & Shen, 2005; see Rains, 2013 for a meta-analytic review). These heightened negative emotions and arousal may then lead to closed communicative intentions, such as perceiving the opposition to be immoral or incompetent (closed mindset), as anger has been shown to result in closed behaviors including attacking and derogating one's opposition (Dillard & Peck, 2001; Smith, 1977).

In sum, we may expect the effects of advocacy style on communicative intentions (regardless of direction) to be driven by the expression of confrontational emotions (e.g. anger), or general negative affect.

The Present Research

To explore the effects of advocacy styles on communicative intentions, we, across two studies, manipulated whether advocates defended their own position (Pro-attitudinal Defense) or attacked their opponent's position (Counter-attitudinal Attack) on an issue related to climate-change. In Study 1 we explored the possibility that Attack (vs. Defend) could lead to more open (vs. closed) communicative behavioral intentions, primarily indexed by information-seeking preferences.

To test between the meta-cognition and emotion hypotheses we also

measured attitude polarization, a range of meta-cognitive variables (including thought favorability, thought confidence, thought conviction, perceived argument quality, perceived effort, and perceived knowledge), and emotions (including anger amongst a range of comparison emotions). Study 2 extended Study 1 by: a) exploring a third hypothesis (motivation hypothesis) as an explanation for our effects in Study 1 and, b) generalizing effects across a range of measures of communicative practice intentions (e.g. choice of interaction partner).

Study 1

Method

Participants. One hundred and thirty-six participants between the ages of 23 and 71 years were recruited via Amazon Mechanical Turk ($M = 41.40$, $SD = 11.78$, female = 62). A sample size of one hundred and twenty-two is sufficient to detect medium effects ($W = .30$) with 80% power at $\alpha = .05$ in our primary analysis, which was a multinomial logistic regression (Faul et al., 2009); we oversampled to allow for exclusions.

Materials and procedure. Participants completed the survey online via Qualtrics (Qualtrics, 2016). They first indicated their Time 1 attitude and attitude confidence towards a range of issues, including a carbon emissions policy, migrant intake, and attitudes towards consumer products. A range of issues was included to reduce participant suspicion about the aims of the study. The issue of primary interest in this study was a carbon emissions policy. Pilot testing ($N = 321$, $M_{\text{age}} = 19.67$, $SD = 4.17$, 17 – 55 years, 240 female) indicated that roughly equal numbers of participants should be on either side of this issue (pilot test numbers: $n_{\text{Pro}} = 155$; $n_{\text{Anti}} = 121$).

Attitudes. Participants were asked how favorable they felt towards a hypothetical carbon emissions policy which was modelled on energy-saving recommendations in the US (Department of Energy, 2018; 1 = *extremely unfavorable*, 7 = *extremely favorable*; $M = 3.23$,

$SD = 2.15$; see Supplementary Materials for full instructions). Those who responded “neutral” (scale point = 4) were re-directed to a different task ($n = 11$), leaving one hundred and twenty-five participants to complete the study²³.

Attitude confidence. Confidence held in this attitude was measured using a single-item (“How much confidence do you have in your attitude towards this policy? 1 = *not at all*, 9 = *very much*; adapted from Tormala & Petty, 2002; $M = 7.22$, $SD = 1.96$).

Following Time 1 attitude measures, participants completed the advocacy task. Participants were randomly allocated to one of two conditions: Defense ($n = 60$) or Attack ($n = 65$). In each condition, participants were asked to write a persuasive message to convince someone who disagrees with them about the carbon emissions policy. Those in the Defense condition were asked to write arguments for *why their view is right* while, those in the Attack condition were asked to write arguments for *why their opponent’s view is wrong* (see Supplementary Materials for full details of the advocacy task).

Following the advocacy task, all participants completed a thought-listing task taken from Clark, Wegener, Sawicki, Petty, and Briñol (2013), in which they listed up to eight thoughts they experienced during the advocacy task. This task was used in conjunction with measures below to assess meta-cognitive confidence.

Time 2 attitude extremity and confidence. Participants indicated their attitude ($M = 3.05$, $SD = 2.37$) and confidence placed in their attitude at Time 2 ($M = 7.92$, $SD = 1.64$), using identical measures to those used at Time 1.

Meta-cognitions. Next, to assess the meta-cognition hypothesis, participants completed a range of meta-cognitive measures in random order (see Supplementary Materials

²³Initial stance coding: Pro = 1, Anti = -1. Pro = favorable towards the policy ($n = 46$), participants who responded with scores of +1, +2 or +3 on the 7-point scale. Anti = unfavorable towards the policy ($n = 79$), those who responded with transformed scores of -1, -2, or -3. No effects of initial stance emerged on attitude polarization (see Supplementary Materials). We additionally measured Need-for-Cognition and the Big 5 as part of a larger study, however these variables were not analyzed as they were not the key focus of this work.

for a full list of items). These included thought favorability (ratio of positive to negative thoughts; $M = -.22$, $SD = .76$), thought confidence ($M_{\text{positive thoughts}} = 8.12$, $SD = 1.24$; $M_{\text{negative thoughts}} = 7.87$, $SD = 1.33$, 9-point scales), thought conviction (attribution of thoughts to core beliefs; $M = 6.79$, $SD = 1.89$, 9-point scale), perceived argument quality ($M = 7.25$, $SD = 1.58$, 9-point scales), perceived effort expended in generating arguments ($M = 7.90$, $SD = 1.16$, 9-point scales), and perceived knowledge on the topic ($M = 5.15$, $SD = 1.30$, 7-point scales).

Emotions. Next, to assess the emotion hypothesis, we measured specific emotions (with a focus on anger; $M = 2.17$, $SD = 1.57$, 7-point scales) and also general affective dimensions of arousal (level of psychological arousal during advocacy; $M = 6.17$, $SD = 1.86$, 9-point scales), and valence (general feelings during advocacy; $M = 3.90$, $SD = 1.85$, 9-point scales), to explore affect-general accounts for advocacy effects (see Supplementary Materials for a full list of items).

Communicative intentions. Finally, participants completed the key dependent measure (information-seeking preferences) and two other behavioral intention measures in a randomized order.

Information-seeking preferences. Participant inclinations to seek attitude confirming and disconfirming information was assessed using the following question: “We may have some information related to the thermostat policy you wrote about previously. What kind of information would you like to receive? You may choose more than one option.”. Participants were given three options (to receive no added information, to receive information about reasons supporting implementation of the policy, and to receive information about reasons against implementation of the policy). Participants were asked to answer the question and enter their email address in order to receive the information (adapted from Vargas, Von Hippel, & Petty, 2004). Responses were coded to reflect whether choice options were Time 1

attitude-consistent only ($n = 16.0\%$), attitude-inconsistent only ($n = 16.8\%$), both ($n = 33.6\%$) or neither ($n = 33.6\%$).

Intentions to engage with the opposition. Positive advocacy-related behavioral intentions (i.e. “open” communicative intentions) were assessed using three 9-point items (“To what extent are you willing to [talk to, listen to, negotiate] with someone holding an opposing viewpoint on this topic?”; 1 = *not at all*, 9 = *very much*). Items were averaged to create a composite score ($M = 6.15$, $SD = 2.18$, Cronbach’s $\alpha = .89$).

Intentions to persuade others. Negative advocacy-related behavioral intentions (i.e. “closed” communicative intentions) were assessed using three 9-point items (e.g. “How likely would you be to persuade your family/friends of your own position on this topic?”, taken from Cheatham & Tormala, 2015). A reliable composite of persuasion intentions was created by averaging across the items ($M = 4.83$, $SD = 2.43$, Cronbach’s $\alpha = .91$).

Results

Data screening and cleaning. Distributions of all continuous variables were examined using boxplots. No influential cases ($> 1.5 \times$ interquartile range; defined by SPSS software) were detected, so all cases were retained in the analyses. All participants’ arguments were on topic and adhered to the manipulation.

Advocacy effects on attitude polarization and meta-cognitive confidence. To evaluate the meta-cognition hypothesis, we first examined any differences in attitude polarization and meta-cognition between the two advocacy styles. Contrary to predictions, advocacy style did not influence attitude polarization²⁴, $t(123) = .09$, $p = .93$. We then proceeded to consider how meta-cognitive confidence differed between advocacy styles.

²⁴ Attitude polarization was calculated by factoring a) the extent of attitude change, i.e. the number of scale points moved from Time 1 to Time 2 (EXTREMITY) and b) the direction of attitude change, i.e. whether the Time 2 attitude moved in a positive or negative direction from Time 1 (DIRECTION). For example, a participant with Time 1 attitude (-2) on the unipolar scale and Time 2 attitude (-3) on the same scale would be allocated a polarization score of +1 because they polarized (moved in an even more “negative” direction compared to their initial attitude) by one scale point (POL score = EXTREMITY $\times$ DIRECTION).

Consistent with previous work measuring multiple meta-cognitive measures (see Briñol et al., 2012), some of our meta-cognitive variables were substantially positively correlated with each other ($.31 < r < .57$). In the interest of parsimony, a principal components analysis using direct oblimin rotation was undertaken in order to reduce the number of variables to a smaller number of components, and to calculate component scores capturing meta-cognitive confidence. The analysis revealed a two-component solution explaining 51% of the variance (see Supplementary Materials for full details of component loadings). Scores representing each component were computed using the Regression method and individual scale items used to measure each variable (Component 1 Cronbach's $\alpha = .63$, Component 2 Cronbach's $\alpha = .79$). Higher scores on Component 1 represent more positive (attitude- consistent) thoughts during advocacy (higher thought favorability)²⁵ and higher perceptions of source credibility. Higher scores on Component 2 represent greater meta-cognitive confidence in one's advocacy attempt (as indexed by perceived argument quality, perceived effort, thought conviction, and thought confidence).

The two advocacy styles only differed on Component 1 ($t(123) = 2.23, p = .03$) and not on Component 2 ($t(123) = -1.67, p = .10$), such that Attack expressed more *negative* thoughts during their advocacy attempt ($M = -.21, SD = 1.14$) compared to Defense ($M = .19, SD = .81$). Taken together, the results suggest that the meta-cognition hypothesis (specifically, thought favorability) may account for some effects of advocacy style on communicative intentions.

Advocacy effects on emotion. To evaluate the emotion hypothesis, we proceeded to examine any differences in emotion. As a first step, we considered whether advocacy styles differed based on anger; they did not, $F(1, 123) = .04, p = .85$. We next considered broader emotion mediation possibilities. First, did advocacy style influence the experience of any of the other discrete emotions? No; independent samples t-tests showed no significant

²⁵ A few participants did not list any issue-relevant thoughts, and thus were missing values on the thought favourability index. To minimize the loss of data, missing values were replaced by the mean.

differences between advocacy styles based on any of the comparison emotions measured, including fear, disgust, happiness, pride, or sadness, $|ts| < 1.2$, $ps > .23$. Second, we considered advocacy style effects on broader valence and arousal dimensions of affect. Although advocacy styles did not differ on valence, $t(123) = 1.02$, $p = .31$, they marginally differed on arousal, $t(123) = 2.02$, $p = .05$, with Attack participants expressing slightly higher levels of emotional arousal compared to Defense ($M_{\text{Attack}} = 6.48$, $SD = 1.80$; $M_{\text{Defend}} = 5.83$, $SD = 1.85$). Taken together, the results suggest that the emotion hypothesis (specifically, emotional arousal) may account for some effects of advocacy style on communicative intentions.

Advocacy effects on communicative intentions.

Information-seeking behavior. To examine the effect of advocacy style on information-seeking behavior, a multinomial logistic regression was performed using advocacy style as the predictor variable (Attack = -1, Defense = 1) and attitude-consistent information as the reference category in order to directly test the effects of advocacy style on uncongeniality bias. Results revealed that, compared to Defense advocates, Attack advocates displayed more open information-seeking intentions, as they were 3.25 times [95% CI: 1.08, 9.77] more likely to seek attitude-inconsistent information over attitude-consistent information ($B = 1.18$, $SE = .56$, Wald $\chi^2(1) = 4.41$, $p = .036$)²⁶.

No other effects obtained significance in this model. (i.e. selecting no information compared to attitude-consistent information, or selecting both kinds of information compared to attitude-consistent information).

Intentions to engage and persuade. In addition to information-seeking preferences, we measured intentions to engage with and to persuade opponents as indices of open vs

²⁶ Given that attitude certainty has been shown to predict uncongeniality bias (see Knobloch-Westerwick & Meng, 2009), the same analysis was run with initial stance and attitude confidence at Time 2. In the model containing the covariates the relationship between advocacy style and information-seeking continued to hold (odds ratio = 3.78 [95% CI: 1.22, 11.74], ($B = 1.33$, $SE = .58$, Wald $\chi^2(1) = 5.30$, $p = .02$). This relationship was not moderated by initial stance ($B = .21$, $SE = .28$, Wald $\chi^2(1) = .56$, $p = .45$) or attitude confidence at Time 2 ($B = .64$, $SE = .37$, Wald $\chi^2(1) = 3.04$, $p = .08$).

closed communicative intentions. However, these two variables were substantially positively correlated [$r(125) = .52$], indicating perhaps a general behavioral inclination to interact with rather than disengage from opponents. To explore the possibility that advocacy influences communicative intentions, we regressed a behavioral intentions composite (average of intention to engage and persuade) onto a contrast-coded condition variable. There was no evidence that advocacy style directly influenced communicative intentions, $\beta = -.18, p = .32$. It is possible, nevertheless, that despite no overall effect of advocacy style on behavioral intentions, that advocacy style partially influences intentions via uncongeniality bias in information-seeking (O'Rourke & Mackinnon, 2016). Given the complexities of formally combining multinomial logistic and linear models within standard statistical mediation models, we instead adopted an informal regression-based approach to considering this question. We regressed the behavioral intentions composite onto a contrast coded condition variable and three orthogonal contrast coded variables capturing variation in information-seeking preferences. Given that the results of multinomial regression above suggested that inconsistent versus consistent information preferences differed by advocacy style and that consistent did not significantly differ from both or none preferences, we generated a set of orthogonal contrasts that focused on capturing the inconsistent versus other information preferences.

Results revealed that preference for inconsistent over other kinds of information (3 -1 -1 -1) significantly predicted behavioral intentions, $\beta = .30, p = .003$. Preferences for consistent over both or none (0 2 -1 -1), $\beta = -.20, p = .11$ was not significant. However, a significant preference for no information emerged over a preference for both kinds of information (0 0 1 -1), $\beta = -.25, p = .009$. Although no formal test of mediation is available for this combination of logistic and linear models, taken together these results suggest that Attack mindsets promote uncongeniality bias in information-seeking behavior and that such a

bias for attitude-inconsistent information is associated with intention to interact with one's attitudinal opponents.

Using meta-cognitive confidence and emotions to predict communicative intentions (meta-cognition and emotion hypotheses). Given that advocacy styles differed on thought favorability, we proceeded to examine if this drove the relationship between advocacy style on information-seeking preferences. Interestingly, thought favorability was not consequential in predicting uncongeniality bias as indicated by a multinomial regression model containing advocacy style as the independent variable and Component 1 as a covariate ($B = .31$, $SE = .30$, $\text{Wald}\chi^2(1) = 1.05$, $p = .30$). The relationship between advocacy style and uncongeniality bias remained significant ($B = 1.07$, $SE = .57$, $\text{Wald}\chi^2(1) = 3.53$, $p = .06$), with no other effects emerging. This finding does not provide evidence in support of the meta-cognition hypothesis.

Further, given that, advocacy styles differed on levels of emotional arousal, we considered whether confrontational emotions accounted for the effect of advocacy style on information-seeking. Adding the arousal composite to the multinomial logistic regression reported above did not account for advocacy style effects on information-seeking behavior (advocacy style effect on attitude-consistent versus inconsistent preferences remained significant: $B = 1.15$, $SE = .57$, $\text{Wald}\chi^2(1) = 4.06$, $p = .04$). This finding does not provide evidence in support of the emotion hypothesis.

Taken together, although Attack increases preference for attitude-inconsistent information, it does not do so via meta-cognitive confidence, such as, by changing thought favorability, or by modulating specific emotional or broader affective experience. Thus, the evidence fails to support both the meta-cognition and emotion hypotheses.

Discussion

To summarize, the results show that Attack mindsets lead to a preference for attitude-inconsistent over consistent information, relative to Defense mindsets. However, this effect is not accounted for by the meta-cognition hypothesis; indeed, advocacy style did not systematically influence information-seeking via meta-cognitive factors. Additionally, we find no support for the emotion hypothesis; the effect is not accounted for by differences in specific confrontational emotions or broader affective factors.

Although we take this as initial evidence of the role of Attack advocacy mindsets in fostering uncongeniality bias, our measures of behavioral intentions were somewhat ambiguous and so we cannot definitively say whether uncongeniality bias is associated with broader ‘open’ communicative practice. It is possible that those in an Attack mindset seek attitude-inconsistent information not in order to improve accuracy in their knowledge of an issue, but to garner additional ‘ammunition’ in order to better out-argue their opponents. In Study 1, our measures of engagement and persuasion intentions were substantially positively correlated, so we could not disentangle these possibilities. In Study 2 we turn our attention to more fully integrating information-seeking within a suite of communication intentions and try to ascertain what drives Attack advocates’ preferences for attitude-inconsistent information.

The results of Study 1 corroborate previous work demonstrating the effects of confrontational mindsets on the desire to overcome congeniality bias (see Kim, 2016; Wiltermuth & Tiedens, 2011; Young et al., 2011). As noted, a possible explanation for this effect is that an Attack mindset increases advocates’ motivation to validate their opinion even further and to garner ammunition against the opposition during future advocacy attempts. Alternatively, it is possible that the Attack mindset increases advocates’ motivation to revise their attitudinal beliefs, by highlighting the utility to be gained by being receptive to a novel perspective. These distinct motivations are termed *defense* motivation and *accuracy*

motivation respectively (Chaiken et al., 1989). In the context of our framework, we suggest that accuracy motivations are more indicative of open communicative intentions and that defense motivations are indicative of closed intentions. In Study 2 we seek to clarify whether Attack mindset effects on uncongeniality are driven by a third hypothesis: the motivation hypothesis. That is, we seek to examine if the effects of advocacy style on open communicative intentions is driven by open, accuracy-based motives or more closed, defense-based motives.

Defense motivation typically manifests as congeniality bias (Hart et al., 2009; Nickerson, 1998) as opposed to uncongeniality bias, given that people tend to be highly motivated to safe-guard their own opinions (Festinger, 1962). However, the Attack mindset may potentially promote uncongeniality bias because it presents advocates with the opportunity to gain ammunition against those who disagree with them (Young et al., 2011), by making salient the incongruence between the opposition's attitude and one's own (Brambilla et al., 2013). That is, advocates may override congeniality bias perhaps in the pursuit of better defending their own position in the future.

On the contrary, accuracy motivation tends to induce uncongeniality bias, especially if seeking attitude-inconsistent information is related to goal accomplishment (see Hart et al., 2009 for a review). Given that the Attack advocacy style draws attention to opposing views, advocates in this mindset may feel that they stand to gain useful issue-specific knowledge by expressing openness to opposing views. That is, *information utility* (the usefulness of information in aiding with future decision-making; see Hart et al., 2009) may induce accuracy motivation in the Attack condition. Heightened accuracy motivation may thus lead to positive downstream consequences such as uncongeniality bias, as indexed by seeking attitude-inconsistent information, or choosing interaction partners holding opposing views.

In sum, we may expect the effects of advocacy style on communicative intentions to be driven by the expression of accuracy versus defense motivations.

Study 2

In addition to exploring the role of accuracy and defense motivations in advocacy effects on information-seeking, in Study 2 we also sought to expand the scope of communicative practice dependent measures. Seeking attitude-inconsistent information is one manifestation of uncongeniality bias; another is seeking attitude-inconsistent interaction partners. Both of these communicative practices demonstrate a desire to actively learn about the opposition whether it be offline or in-person, and thus classify as open communicative intentions.

In addition to uncongeniality bias, in Study 2 we also sought to assess other factors associated with open communication and interaction. We assess “open mindset” by measuring perceptions of and receptivity to opponents. Receptivity to opponents is captured by willingness to listen to and engage with the opposition’s views, while person perception is captured by stereotype content used to describe the opposition group, such as, warmth, competence, and morality (Fiske, Cuddy, & Glick, 2007; Goodwin, Piazza, & Rozin, 2014). Stereotype content is important because it carries downstream consequences on affective and behavioral reactions towards outgroups. For example, positively perceiving the opposition as high in warmth leads to approach behaviors, while negatively perceiving the opposition as low in competence and morality leads to avoidance or even attacking the outgroup (see Fiske, 2015; Yzerbyt, 2016 for reviews).

Thus, the two key aims of Study 2 were to: a) broaden the scope of open intentions to include information-seeking, partner selection, opponent perception and receptivity to opponents’ views, and b) examine whether effects of advocacy on such measures are driven by accuracy or defense motives (motivation hypothesis).

Method

Participants. One-hundred and ninety-two participants between the ages of 15 and 74 years were recruited via Amazon Mechanical Turk ($M = 35.80$, $SD = 10.97$, female = 84). Again, this sample size was sufficient to detect medium effects ($W = .30$) in multinomial logistic regressions, and was also sufficient to detect small-medium effects ($f^2 = .08$) with 80% power, in a linear regression containing five predictors (Faul et al., 2009). We oversampled to allow for exclusions (e.g. neutral participants).

Materials and Procedure. Participants completed the survey online. The measures used in this study were identical to Study 1, with a few exceptions detailed below.

Attitudes. Similar to Study 1, participants first indicated their attitude towards the carbon emissions policy ($M = 3.33$, $SD = 2.03$) and how much confidence they placed in their attitude ($M = 7.09$, $SD = 1.89$). Similar to Study 1, more participants were classified as Anti ($n = 111$) compared to Pro ($n = 67$). Those who responded as “neutral” to the attitude issue ($n = 14$) were excluded from the study.

Participants were then directed to the advocacy task which was identical to Study 1. As in Study 1, participants were randomly allocated to generate arguments in either the Defense ($n = 86$) or Attack ($n = 92$) advocacy style. Similar to Study 1, participants then completed a thought-listing task, and again indicated their attitude position ($M = 3.27$, $SD = 2.27$) and attitude confidence at Time 2 ($M = 7.64$, $SD = 1.81$).

Meta-cognitions. To validate the results of the meta-cognition hypothesis in Study 1, participants responded to the same meta-cognitive measures as in Study 1 in a randomized fashion (confidence in positive thoughts: $M = 7.76$, $SD = 1.60$; confidence in negative thoughts: $M = 7.30$, $SD = 2.09$; thought favorability: $M = -.14$, $SD = .69$; thought conviction:

$M = 6.74$, $SD = 2.09$; perceived argument quality: $M = 6.92$, $SD = 1.60$; perceived effort: $M = 7.69$, $SD = 1.39$; perceived knowledge : $M = 4.78$, $SD = 1.90$).

Emotions. To validate the results of the emotion hypothesis, we measured general affective dimensions of arousal and valence, given the differences between Attack and Defense on emotional arousal in Study 1. We did not measure specific emotions given the absence of any differences in Study 1. Participants responded to items assessing emotional valence ($M = 3.99$, $SD = 1.87$, $r = .80$) and emotional arousal ($M = 6.25$, $SD = 1.77$, $r = .50$) in a randomized fashion.

Next, participants completed novel measures of motivations, to test the motivation hypothesis in addition to the meta-cognition and emotion hypotheses.

Motivations. Participants' motivations to engage with the opposition were assessed using six 9-point items (where 1 = *not at all*, 9 = *very much*).

Defense motivation. Defense motivation was assessed using three items ("To what extent are you willing to engage with someone holding an opposing viewpoint on this topic for the purpose of [winning an argument, proving that you are right, showing your opponent why they wrong]", $M = 4.80$, $SD = 2.14$, Cronbach's $\alpha = .88$).

Accuracy motivation. Accuracy motivation was also assessed using three items ("To what extent are you willing to engage with someone holding an opposing viewpoint on this topic for the purpose of [learning more, having an accurate representation of the world, obtaining a more balanced view on this topic]", $M = 6.21$, $SD = 2.05$, Cronbach's $\alpha = .89$).

Communicative intentions. Participants then completed the following behavioral measures presented in a random order: information-seeking preferences, choice of interaction partner, receptiveness to opposing views, and judgements of the opposition.

Information-seeking preferences. Participants completed the same information-seeking measure as in Study 1. Responses were coded to reflect whether choice options were

Time 1 attitude-consistent only ($n = 30.5\%$), attitude-inconsistent only ($n = 23.7\%$), both ($n = 16.9\%$), or neither ($n = 28.8\%$).

Choice of interaction partner. In addition to information-seeking preferences, we assessed advocates' likelihood of choosing to interact with someone holding an opposing view by asking who they would chose to interact with online [options: partner who holds the same opinion as you (52.6%), partner who holds the opposite opinion to you (47.4%; adapted from Rios, Goldberg, & Totton, 2018, see Supplementary Materials for full instructions)].

Open mindset. This was measured by assessing receptivity to opposing views, and perceptions of the opposition (see Supplementary Materials for all items). Participant's receptiveness to opposing views was assessed using three 9-point scales, (1 = *not at all*, 9 = *extremely*, $M = 5.57$, $SD = 1.76$, Cronbach's $\alpha = .84$). To assess judgements of the opposition, participants rated perceptions of someone who disagrees with them on the attitude issue by answering nine questions on 7-point Likert scales assessing perceptions of competence, warmth, and morality (taken from Goodwin, Piazza, & Rozin, 2014). Average scores on each of these dimensions were highly correlated ($.68 > r > .82$) and thus combined into a single index representing person perception ($M = 4.55$, $SD = 1.88$, Cronbach's $\alpha = .93$).

Previous advocacy experience. Finally, given that past communicative intentions may influence future communicative intentions (McGuire, 1961), we assessed advocates' previous experience of defending and attacking on similar attitude issues to be used as covariates. These were measured using four five-point scales (see Supplementary Materials for exact items). Given the high correlations between items, two composites representing past frequency of Attack ($M = 2.23$, $SD = .99$, $r = .75$) and defense ($M = 2.52$, $SD = 1.11$, $r = .85$) were created.

Results

Similar to Study 1, boxplots of continuous variables revealed no influential cases; thus, all cases were retained in the analyses.

Advocacy effects on attitude polarization and meta-cognitive confidence.

Consistent with the results of Study 1, Attack and Defense did not vary on attitude polarization ($t(176) = .89, p = .37$). Also consistent with Study 1, a principal components analysis revealed a two-component solution explaining 48% of the variance in meta-cognitive confidence (see Supplementary Materials for full details of component loadings). The two components were very similar to Study 1 (Component 1: thought favorability, source credibility; Component 2: confidence in negative and positive thoughts, perceived argument quality, perceived effort, thought conviction). Again, the two advocacy styles only differed on Component 1, ($t(176) = -3.50, p = .001$), and not on Component 2, $t(176) = 1.06, p = .29$), such that Attack displayed lower thought favorability ($M = -.24, SD = .87$) compared to Defense ($M = .27, SD = 1.07$).

Advocacy effects on emotion. Similar to Study 1, advocacy styles did not differ on emotional valence, $t(176) = -.46, p = .65$. However, they also did not differ on levels of emotional arousal, $t(176) = .28, p = .78$, which was unlike in Study 1 in which Attack participants demonstrated marginally higher levels of emotional arousal compared to Defense participants.

Advocacy effects on accuracy and defense motivations.

To assess effects of advocacy style on accuracy and defense motivations, each motivation was regressed onto a contrast-coded condition variable (Defense = 1; Attack = -1), along with the measures of past Attack and Defense behavior as a covariate. Also, given that accuracy and defense motivations were moderately correlated with each other [$r(178) = .35$], in each of the regressions, the alternative motivation was also included as a covariate. In

these regression contexts, advocacy style significantly predicted accuracy motivation, such that Attack advocates reported greater accuracy motivation than Defense advocates, $\beta = -.15$, $p = .03$; advocacy style did not, however, predict defense motivation, $\beta = .04$, $p = .55$.

Taken together, the results suggest that differences accuracy motivation (motivation hypothesis) may potentially drive differences in communicative intentions between advocacy styles.

Advocacy effects on communicative intentions.

Information-seeking behavior. Unlike in Study 1, we failed to find an overall effect of advocacy style on information-seeking behavior in a multinomial logistic regression ($B = -.13$, $SE = .41$, Wald $\chi^2(1) = .10$, $p = .76$)²⁷. Interestingly, despite the advocacy styles varying on accuracy motivation, adding a studentized residual of accuracy motivation without defense, did not significantly alter the results of the regression ($B = -.19$, $SE = .42$, Wald $\chi^2(1) = .11$, $p = .66$), with accuracy motivation failing to predict information-seeking preferences on its own ($B = .19$, $SE = .22$, Wald $\chi^2(1) = .74$, $p = .39$).

Choice of interaction partner. Next, we conducted a logistic regression to examine the effects of advocacy style (Attack = -1, Defend = 1) on choice of interaction partner (same opinion as me = 0, different opinion = 1). Similar to information-seeking preferences, results revealed that advocacy style did not significantly predict choice of interaction partner ($B = .14$, $SE = .30$, Wald $\chi^2(1) = .23$, $p = .63$)²⁸. Notably, while adding the studentized residual of accuracy motivation did not significantly alter the overall relationship ($B = -.10$, $SE = .33$, Wald $\chi^2(1) = .08$, $p = .78$), accuracy motivation significantly predicted choice of interaction partner on its own ($B = .87$, $SE = .19$, Wald $\chi^2(1) = 21.35$, $p < .001$), such that for every one

²⁷ Despite advocacy styles varying on thought favorability, similar to Study 1, we find no significant effect of thought favorability on information-seeking preferences ($B = -.20$, $SE = .32$, Wald $\chi^2(1) = .38$, $p = .54$), thus providing no evidence for the meta-cognition hypothesis.

²⁸ While Attack and Defense differed on thought favourability, this was not consequential in predicting choice of interaction partner in a separate multinomial logistic regression containing advocacy style and thought favourability as predictors ($B = -.20$, $SE = .23$, Wald $\chi^2(1) = .70$, $p = .40$). Again, this provides no evidence in support of the meta-cognition hypothesis.

unit increase in the studentized residual of accuracy motivation, we may expect a 2.39 times higher likelihood [95% CI: 1.65, 3.47] of choosing an interaction partner with a different as opposed to similar opinion. This finding provides partial support for the motivation hypothesis, and suggests a potential indirect effect of advocacy style on choice of interaction partner via accuracy motivation.

Open mindset. Finally, we examined the effects of advocacy style on open mindset. A principal components analysis using oblimin rotation indicated that the two measures of open mindset (receptivity to opposing views and person perception) loaded onto a single component explaining 70% of the variance (component loadings: receptivity to opposing views = .84; person perception = .84). Component scores (Regression Method) generated by the principal components analysis were used to represent open mindset. Higher scores reflect more positive perceptions of the opposition. Similar to the effects of advocacy on choice of interaction partner, a linear regression revealed no significant effect of advocacy style on open mindset ($R^2 = .00$, $F(1, 176) = .47$, $p = .50$)²⁹.

However, similar to choice of interaction partner, conducting a linear regression with advocacy style and accuracy motivation revealed a significant effect of accuracy motivation on open mindset ($\beta = .37$, $p < .001$), such that increasing accuracy motivation resulted in greater open mindset ($R^2 = .14$, $F(2, 175) = 13.99$, $p < .001$). This finding also partially supports the motivation hypothesis, and suggests some indirect effects of advocacy style on open mindset, via accuracy motivation.

Using accuracy motivation to predict communicative intentions (motivation hypothesis). The results of the previous regression analyses are consistent with the possibility that accuracy motivation (motivation hypothesis) may account for advocacy style effects on particular communicative intentions, such as choice of interaction partner and open

²⁹A linear regression containing advocacy style and thought favorability ($R^2 = .00$, $F(2, 175) = .25$, $p = .78$), revealed no significant effect of thought favorability ($\beta = .02$, $p = .87$) on open mindset either, despite condition-level differences, again providing no evidence for the meta-cognition hypothesis.

mindset. To formally explore accuracy motivation as a potential mediator, we conducted mediation analyses using PROCESS (Hayes, 2013), in which accuracy and defense motivation were entered as simultaneous mediators into two mediation models containing advocacy style as the independent variable, with choice of interaction partner and component scores from the open mindset measure as dependent variables, and past experience defending and attacking as covariates. Models and parameter estimates are presented in Figure 6-1 (note, advocacy style coded as Defense = 1, Attack = -1).

Choice of interaction partner (uncongeniality bias). A mediation analysis using simultaneous mediators, 5000 bootstrapped samples and bias-corrected 95% confidence intervals (see Preacher & Hayes, 2008) showed that although there was no direct effect of advocacy style on choice of interaction partner [DE = .03, SE=.17, 95% CI = -.30, .36]³⁰, there was a significant indirect effect via accuracy motivation [IE= -.14, SE =.08, 95% CI = -.33, -.01], but not via defense motivation [IE = .00, SE=.02, 95% CI = -.03, .04], controlling for past frequency of attacking [on accuracy motivation: B =.13, SE =.10, p =.22; on defense motivation: B =.49, SE =.09, p < .001], and past frequency of defending one's position [on accuracy motivation: B =.19, SE =.09, p =.04; on defense motivation: B =.03, SE =.08, p =.70] (see Figure 1, Panel A). This finding supports the motivation hypothesis, and suggests that Attack increases uncongeniality bias via increasing accuracy, not defense, motivations.

Open mindset. Similarly, although there was no direct effect of advocacy style on open mindset [DE= .00, SE=.07, 95% CI = -.14, .14], there was a significant indirect effect via accuracy motivation [IE= -.06, SE=.03, 95% CI = -.14, -.01], but not via defense motivation [IE= .00, SE=.01, 95% CI = -.01, .02], controlling for past frequency of attacking and defending one's position (see Figure 6-1, Panel B). Again, this finding supports the

³⁰ DE = direct effect, IE = indirect effect. All effects are reported as standardized regression coefficients.

motivation hypothesis, and suggests that Attack increases open mindsets via increasing accuracy motivation, not defense motivation.

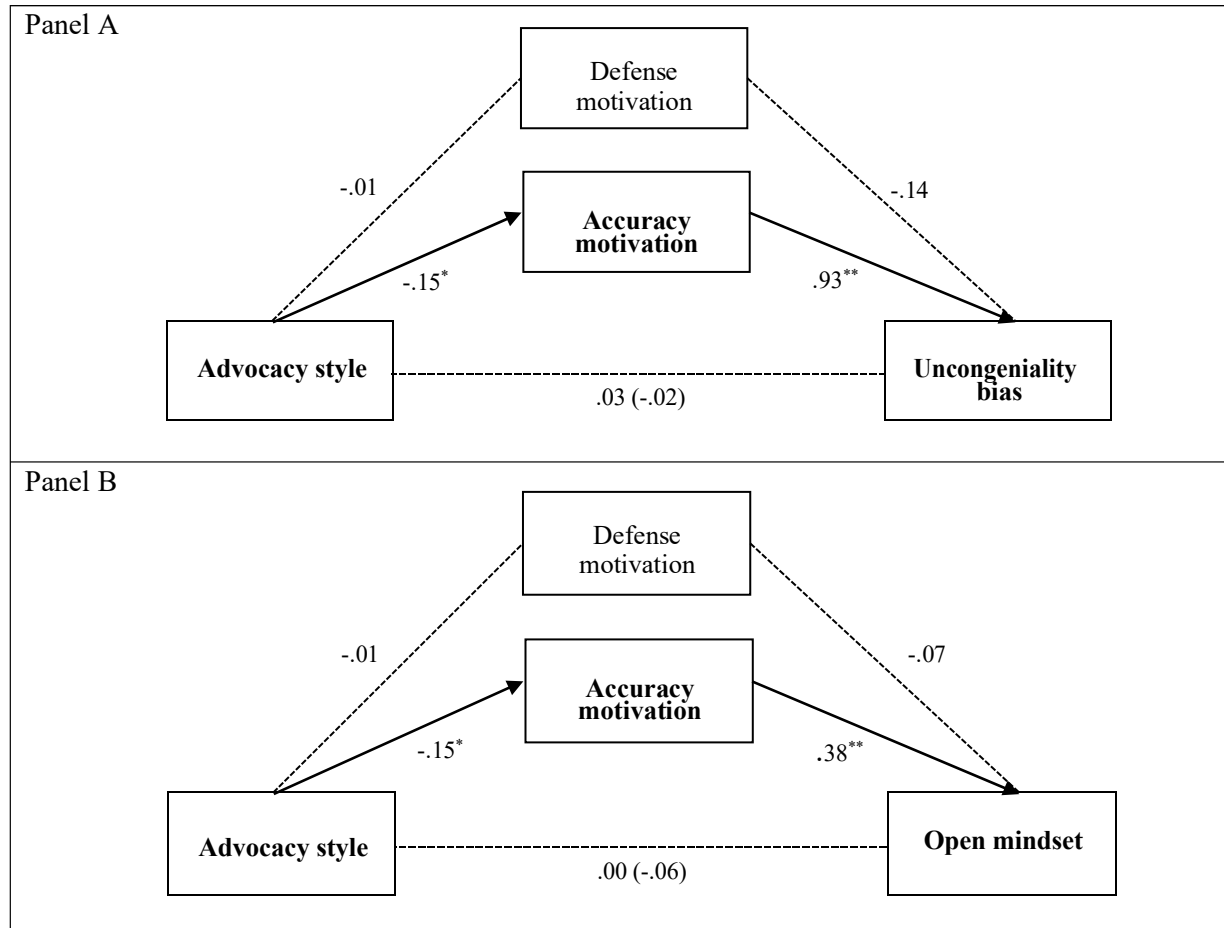


Figure 6-1. Indirect effect of advocacy style on uncongeniality bias via accuracy motivation is depicted in Panel A, and effects on open mindset in Panel B.

Advocacy style coded Defense = 1; Attack = -1. The standardized coefficient within brackets represents the total effect of advocacy style on the dependent variable with covariates included in the model, but without the mediators. Each of the models controls for past frequency of Attack and Defense, but these are omitted from the diagram for clarity of presentation.

Note. * $p < .05$ ** $p < .01$.

General Discussion

The current research aimed to investigate the effects of advocacy style on communicative intentions. Specifically, we were interested in the effects of Counter-attitudinal Attack (attacking the opponent's position) versus Pro-attitudinal Defense (defending one's own position) on advocates' attitudes and subsequent intentions to engage in "open" versus "closed" communicative practices. Results revealed that, paradoxically, compared to Defense, Attack led to more open communicative intentions, as captured by uncongeniality bias and open mindset. Importantly, this effect was mediated by accuracy motivation (motivation hypothesis) but not changes in attitude polarization, meta-cognitive confidence (meta-cognition hypothesis), confrontational emotions (emotion hypothesis), or defense motivation.

Summary

This research was the first to explore Attack vs Defense in the context of pro-attitudinal advocacy, while bringing an interdisciplinary approach guided by social psychology, political psychology, and the communications literature. We proposed three hypotheses through which attitudes and/or communicative intentions may be influenced as a result of Attack vs Defense: meta-cognition hypothesis, emotion hypothesis, and the motivation hypothesis. While we found partial support for both meta-cognition and emotion hypotheses due to condition-level differences between the two advocacy styles (Study 1 and Study 2), neither of these were consequential in predicting communicative intentions. Rather, we find support for the motivation hypothesis, such that accuracy (but not defense) motivation, is consequential in predicting choice of interaction partner and open mindset, due to higher levels of accuracy motivation expressed by Attack compared to Defense advocates (Study 2). These findings collectively suggest that under certain circumstances, Attack advocates may display greater willingness to engage with the opposition (in the form of

interaction partners or information), not for the purpose of gaining greater leverage or to confirm existing worldviews, but, rather, for the purpose of holding more accurate worldviews. This is supported by the fact that accuracy - but not defense - motivation indirectly mediated the effect of advocacy style on: a) uncongeniality bias (as measured by choice of interaction partner holding dissimilar views), and b) increased receptivity to and positive perceptions of the opposition (as measured by the open mindset measure).

Interestingly, we find no variation between advocacy styles on attitude polarization, but rather a direct effect of motivations on behavior. This finding is consistent with previous work highlighting that attitudes towards controversial issues (such as climate-change used in the current study) are difficult change (Krosnick & Petty, 1995), but that behavioral intentions may change even if attitudes do not, in the context of pro-attitudinal advocacy (Abeywickrama et al., 2020).

Is It Good to Attack?

Although they may appear counter-intuitive, the current findings are consistent with previous work illustrating that people primed with confrontational mindsets (i.e. those induced with anger) are more likely to overcome congeniality bias by seeking or generating attitude-inconsistent information or by expressing interest in alternative viewpoints (Kim, 2016; Wiltermuth & Tiedens, 2011; Young et al., 2011). However, in our study, Attack advocates overcame congeniality bias not because of anger or defensiveness, but through a desire to hold more accurate world-views.

This critical finding raises the question: *why* does attacking an opponent motivate one to be more accurate? One possibility is that Attack advocacy induces a temporary perspective-taking state or *change of attentional focus* owing to its emphasis on the opponent's attitude as opposed to one's own. By drawing attention to the opposition's perspective, Attack advocates may be inclined to evaluate not only their opposition's

arguments but also to scrutinize the merits of their *own* arguments. In other words, encouraging advocates to deliberate upon the opposition's point of view may heighten accuracy motivation by drawing attention to the cogency (or lack thereof) of their own arguments.

Another possibility is that Attack somehow draws attention to the novelty of the topic (carbon emissions policy) and the utility of information to be gained by interacting with the opposition. Drawing attention to the opponent's view may result in advocates questioning their knowledge on the topic, which they would use as a basis for arguing against the opposition. In particular, Attack advocates may feel that they lack issue-specific knowledge (e.g. knowledge of carbon emissions policies) to argue against the opposition, although they may have confidence in their ability to argue on the topic more generally (knowledge of climate-change). This notion is supported by the fact that Attack advocates reported lower thought favorability during advocacy, but expressed a similar level of meta-cognitive confidence as their Defense counterparts.

Notably, accuracy motivation reduces congeniality bias in contexts where people are motivated to make the best choice or in situations where people experience greater uncertainty (Fischer & Greitemeyer, 2010). For instance, undermining confidence in people's decisions by framing choice options in terms of losses reduces congeniality bias in subsequent information searching (Fischer, Jonas, Frey, & Kastenmüller, 2008). In our study, the Attack condition potentially undermined advocates' confidence in their issue-specific knowledge of the topic, the merits of their arguments, or both factors, leading to greater accuracy motivation. This heightened accuracy motivation subsequently led to increased willingness to engage with the opposition.

Further, the attitude issue selected, although an issue raised by the Department of Energy, was likely unfamiliar to many. Thus, advocates may have felt that they had much to

gain by selecting novel, attitude-inconsistent information. Indeed, novel information draws people's attention by encouraging them to "stop and think" (Shoemaker, 1996). Moreover, corroborating our findings, novel information stimulates people to choose news articles even if they contain counter-intuitive information (Lee, 2008). Therefore, Attack may draw attention to the information utility to be gained by engaging with opposing viewpoints, stimulating uncongeniality bias and receptivity to those holding opposing views.

Indeed, previous work shows that information with high utility (information pertinent to achieving a goal or outcome) is more likely to be selected even if that information is attitude-inconsistent (Hart et al., 2009). In other words, information utility can *override* congeniality bias (Knobloch-Westerwick & Kleinman, 2012). In our study, the prospect of information utility may have stimulated a desire for advocates to learn more about the issue by selecting information and also interaction partners that have new information to contribute.

Limitations and Future Directions

The limitations of the current research can be summarized across two key themes: a) replicability and generalizability concerns, and b) alternative explanations. Each of these will be examined in-turn.

Replicability & generalizability.

Replicability of results. Our findings for information-seeking preferences in Study 1 (Attack displaying greater preferences for attitude-inconsistent vs attitude-consistent information), did not replicate in Study 2. Further, accuracy motivation failed to predict information-seeking preferences (preference for attitude-inconsistent vs attitude-consistent information only) in Study 2. It could be argued that uncongeniality bias can be construed more broadly, such that if one intends to hold more accurate worldviews, one may wish to

select *both* kinds of information (not only attitude-inconsistent information) compared to attitude-consistent information only.

However, a post-hoc multinomial regression analysis in Study 2 reveals that advocacy style does not predict the selection of both kinds of information compared to attitude-consistent information only ($B = -.42$, $SE = .47$, Wald $\chi^2(1) = .78$, $p = .37$). However, we do find a significant effect of the studentized residual of accuracy motivation on information-seeking ($B = .58$, $SE = .26$, Wald $\chi^2(1) = 5.30$, $p = .02$), such that for every one unit increase in the studentized residual of accuracy motivation, we may expect a 1.8 times higher likelihood [95% CI: 1.09, 2.97] of an advocate choosing both kinds of information compared to attitude-consistent information only³¹. Despite the failure to replicate the overall finding in Study 2, these results suggest that the desire to hold more accurate worldviews may nevertheless account for indirect effects of advocacy style on uncongeniality bias, as conceptualized in different forms. Our finding of accuracy, but not defense motivation significantly mediating Attack advocates' choice of interaction partner with dissimilar views, provides further evidence for this claim.

Generalizability across attitude issues. The generalizability of our findings to other attitude issues must be investigated by future research, given that we studied a single attitude issue (climate-change). Different attitude issues may elicit different reactions from advocates, and subsequently differentially impact depolarization and communicative intentions. For instance, previous work in self-persuasion (group polarization) shows that depolarization is less likely to occur for controversial attitude issues such as capital punishment, but more likely to occur for less morally convicted issues, such as, choosing product alternatives, suggesting that the extent of depolarization (and possibly communicative intentions) is

³¹ Although Attack displayed significantly lower thought favorability compared to Defense, this was not consequential in predicting the selection of both kinds of information compared to attitude-consistent information only ($B = .01$, $SE = .35$, Wald $\chi^2(1) = .00$, $p = .98$), again suggesting support for the motivation hypothesis but not the meta-cognition hypothesis.

proportional to investment in the issue (Stoner, 1968; Vinokur & Burnstein, 1978). We saw movement in information-seeking preferences and communicative intentions for an arguably highly morally and affectively-laden issue related to climate-change, which is promising. Despite only employing one attitude issue, we followed similar preliminary work in communication research exploring effects of arguing on meta-cognition (Walter & Cohen, 2019), to ensure the validity of our results by: a) undertaking supplementary analyses to ensure that other factors, such as, initial stance did not affect the results, b) using the same measures across both studies with minimal changes, to ensure replicability, and c) re-running key analyses on both studies to rule out alternative explanations for our findings.

Generalizability to free vs incentivized advocacy. There is a question around the external validity of our results when people advocate freely versus when offered some incentive. In the current research advocates were required to complete the advocacy task in exchange for a monetary incentive. Future research should examine the consequences of pro-attitudinal advocacy when advocacy arises organically and without prompting. Some evidence from counter-attitudinal advocacy suggests that people are more likely to depolarize (and thus become more open to the opposition) under free versus forced advocacy contexts - when external reinforcement is minimal, advocates have no choice but to change their opinion in order to reconcile their cognitive dissonance (Miller & Wozniak, 2001). In the context of pro-attitudinal advocacy, it is possible that our effects will be exacerbated during free-advocacy under certain conditions, such as, when meta-cognitive confidence is low. Future research may test this intriguing possibility.

Alternative explanations.

Lack of control condition. One may argue that the inferences which can be drawn based on the current research are limited due to lack of a “true control” condition. That is, any effect between the two experimental conditions may be asymmetrical around a control-

midpoint. In this case, it's not clear as to whether compared to a no-advocacy control, Attack leads to more open practices, or Defense leads to more closed practices, or both. Including a no-advocacy control condition would help clarify this in future research.

Other mechanisms. While the current research explored multiple explanations for our effects, there may be alternative explanations for our findings. For example, one may argue that the current findings can be explained by phenomena other than for motivations. For instance, Attack can be associated with approach mindsets (De Dreu & Gross, 2018), with readiness (Briñol, Petty, & Requero, 2017), or with defensive confidence (Bizer, Larsen, & Petty, 2011). While Attack and Defense displayed some variation in constructs other than for accuracy motivation (e.g. emotional arousal and thought favorability; Study 1), we did not find the relationship between advocacy style and uncongeniality bias to be driven by approach-motivated emotions, such as, anger (Carver & Harmon-Jones, 2009), emotional arousal or constructs indicating defensive confidence, such as, meta-cognitive confidence. Thus, we find little evidence to support the above alternative explanations. Nevertheless, future work may explicitly test these different possibilities to more fully explore the mechanisms underlying the relationship between advocacy style and uncongeniality bias.

Self vs other confound. Some may suggest that our effects are driven by attacking *others*, versus defending the *self*. That is, one may argue that target or focus of advocacy (own position vs other's position) drives the effects, as opposed to style of advocacy (attacking vs defending) only. We present two key reasons for why this is not a major concern in the current research. First, the advocacy task instructions sufficiently highlighted the target of advocacy, or the "other" in both Attack and Defense conditions. That is, advocates were asked to focus on persuading a target with a different viewpoint in both advocacy conditions, as opposed to only those in the Attack condition (see Supplementary Materials for task instructions).

Second, previous work in pro-attitudinal advocacy shows no effects on dependent variables, such as attitude polarization, based on the target of advocacy (Abeywickrama & Laham, 2020a; Briñol et al., 2012). That is, asking advocates to focus on a target possessing a different position to one's own, versus a target possessing the same position to one's own has no effect on attitude change post-advocacy (Abeywickrama & Laham, 2020a). If the "self versus other" target of advocacy account was true, those focusing on persuading a counter-attitudinal target are likely to experience different attitude-change and behavioral outcomes compared to those attempting to convince a pro-attitudinal target. Taken together, this suggests that target of advocacy is likely less important in the current context, but may warrant further attention in future advocacy research.

Conclusions

Taken together, our findings suggest that Attack mindsets paradoxically yield more positive advocacy-related intentions, such as, a preference for attitude-inconsistent information, and increased willingness to interact with the opposition than do Defense mindsets. These findings carry important implications for reducing closed communicative intentions resulting from political advocacy, and thus bridging political divides. Our findings suggest that encouraging advocates to shift their attentional focus to opposing viewpoints may subsequently lead to more civil discourse between people from different attitudinal groups.

Acknowledgements

The authors wish to thank Gajan Thiyagarajah for his useful comments in improving the clarity of the manuscript.

Funding details

This research was supported by an Australian Government Research Training Scholarship awarded to the primary author for the completion of their PhD dissertation.

Declaration of interest

The authors declare no conflicts of interest.

Data availability statement

The data that support the findings of this research are openly available on the Open Science Framework at <https://osf.io/4d3sr/>.

Supplementary Materials

Study 1

Method.

Attitude issue. “Carbon dioxide emissions from electricity and gas usage contribute to more than half of America’s total carbon dioxide emissions. These emissions peak particularly during colder times of the year, due to significantly increased usage of heating. The higher the household thermostat temperature, the greater the energy consumed. The US government may therefore pass a law specifying that households set a maximum thermostat temperature of 68 Fahrenheit (20 degrees Celsius) during winter.”

Advocacy task instructions. Pre-amble. “When people try to persuade others, they can do so in different ways. For example, they might give arguments for why they are right; that is, they may support and defend their own point of view. On the other hand, they may give arguments for why their opponent is wrong; that is, they might try to undermine and attack their opponent’s point of view in an attempt to persuade”.

Defense condition. “Imagine that you have to write arguments for someone who is [favorable/unfavorable] towards a government policy specifying a maximum thermostat temperature of 68 Fahrenheit (20 degrees Celsius) during winter. That is, the person you are trying to persuade disagrees with you on this topic. Think of arguments for why your view is right. Focus only on your own attitude; do not explain why your opponent is wrong, but why you are right. That is, persuade your opponent of your own view, by telling them specifically why YOU are RIGHT.”

Attack condition. “Imagine that you have to write arguments for someone who is favorable towards a government policy specifying a maximum thermostat temperature of 68 Fahrenheit (20 degrees Celsius) during winter. That is, the person you are trying to

persuade disagrees with you on this topic. Think of arguments for why your opponent's view is wrong. Focus on your opponent's attitude; do not explain why you are right, but why they are wrong. That is, persuade your opponent of your own view, by telling them specifically why THEY are WRONG."

Meta-cognitions.

Thought favorability. Participants were asked to reflect on the thoughts that they listed earlier and to rate their valence as issue *positive, negative, neutral, or unrelated to the issue* (see Clark et al., 2013). An index of thought favorability was computed by subtracting the number of negative thoughts from positive thoughts, and dividing this by the total number of issue-relevant thoughts $M = -.22, SD = .76$).

Thought confidence. Confidence in each of the thoughts listed previously was assessed using a single item 9-point Likert scale (1 = *not at all*, 9 = *very much*; taken from Tormala & Petty, 2002). For those who listed positive thoughts ($n = 54$), the average confidence in positive thoughts was 8.12 ($SD = 1.24$); for those who listed negative thoughts ($n = 86$) the average confidence in negative thoughts was 7.87 ($SD = 1.33$).

Thought conviction. Attribution of thoughts to participants' core beliefs was assessed using two items ("My thoughts in response to the issue of a maximum thermostat temperature reflect something enduring about who I am[are connected to my core beliefs]", 1 = *not at all*, 9 = *very much*). Responses were averaged to create a reliable composite ($M = 6.79, SD = 1.89, r = .76$).

Perceived argument quality. Participants rated how strong or weak they perceived their own arguments to be on three 9-point items (e.g. "How strong are the arguments that you generated?"; adapted from Briñol, McCaslin, & Petty, 2012). Responses were averaged to create a reliable composite score ($M = 7.25, SD = 1.58, \text{Cronbach's } \alpha = .87$).

Perceived effort. Effort expended in generating arguments was assessed using three 9-point items (e.g. “How much energy did you put into generating your arguments?”, taken from Briñol et al., 2012; $M = 7.90$, $SD = 1.16$, Cronbach’s $\alpha = .77$).

Self-perceived knowledge. Perceived knowledge on carbon emissions was assessed using three 7-point Likert scales (“I have heard of issues related to carbon emissions”, “I have researched issues related to carbon emissions”, “I understand issues related to carbon emissions”, 1 = *strongly disagree*, 7 = *strongly agree*; adapted from Ehret, Van Boven, & Sherman, 2018). Responses were averaged to create a reliable composite of self-perceived knowledge ($M = 5.15$, $SD = 1.30$, Cronbach’s $\alpha = .80$).

Emotions.

Specific emotions. Although our focus was on anger, we measured a range of emotions such as fear, sadness, disgust, happiness and pride for comparison purposes. Each emotion was measured using two 7-point scales (e.g. anger: “How angry[resentful] did you feel while completing the task?”, 1 = *not at all*, 7 = *very much*). Composites was created by averaging across the two items ($.63 < r < .72$; anger: $M = 2.17$, $SD = 1.57$, $r = .72$).

Emotional arousal. Psychological arousal during advocacy was assessed using two 9-point semantic differentials (passive-active, unaroused-aroused; adapted from Berger, 2011). Given the strong correlation ($r = .73$), items were averaged to form a composite score of psychological arousal during advocacy ($M = 6.17$, $SD = 1.85$).

Emotional valence. Participants indicated, in general, how they felt during the task by responding to two 9-point semantic differential scales (pleasant-unpleasant, good-bad; adapted from Briñol, Petty, & Barden, 2007). Items were averaged to form a reliable composite of general emotional valence ($M = 3.90$, $SD = 1.85$, $r = .90$).

Results.

Predicting attitude polarization from advocacy style and initial stance. To examine polarization as a function of advocacy style and initial stance, we conducted a 2(Advocacy style: Defense vs Attack) x 2(Initial stance: Pro vs Anti) between-subjects ANOVA with the polarization index as the dependent variable. There was no significant effect of advocacy style $F(1, 121) = .26, p = .61$, or initial stance $F(1, 121) = 2.61, p = .11$ on polarization. There was also no interaction between advocacy style and initial stance, $F(1, 121) = 1.57, p = .21$. This eliminates Hypothesis 1, which suggested that advocates in the Attack condition will exhibit greater attitude polarization compared to those in the Defense condition. We repeated the above analysis with attitude confidence change as the dependent variable, however, there was no significant effect of advocacy style on attitude confidence change, $F(1, 121) = .10, p = .76$.

Principal components analysis. Given that not all participants indicated positive and/or negative thoughts, and that conducting analysis listwise or pairwise would result in the loss of data, for each variable missing values were replaced with the variable mean. Component loadings from the pattern matrix indicated that thought favorability (.79) and perceived source credibility (.65) loaded onto one component (Component 1) while perceived argument quality (.84), perceived effort (.70), thought conviction (.70), confidence in positive thoughts (.52) and confidence in negative thoughts (.58) strongly loaded onto a separate component (Component 2).

Study 2

Method.

Selection of interaction partner instructions. “Imagine that we gave you the opportunity to interact online with a person who is roughly the same age, sexual orientation and political orientation as you. From these interaction partners, you can choose to chat with

either someone who holds the same view as you on the issue related to carbon emissions policy, or someone who holds the opposite view to you on this issue.

Open mindset items.

Receptivity to opposing views.

“How receptive would you be to the views of someone who holds the *opposite* opinion to you on this issue? “How receptive would you be to the arguments of someone who *disagrees* with you on this issue?” “How accepting would you be of the views of someone who disagrees with you on this issue?”

Person perception instructions and items. Think of the kind of person who holds the opposite opinion to you on the issue of climate-change. Answer the following questions related to how you perceive that person.

“To what extent do you consider this person to be [skilled, competent, intelligent, likeable, friendly, warm, honest, sincere, trustworthy]?”

Previous advocacy experience.

Defense. “How frequently have you had to defend your position on issues related to climate-change? “In the past, how frequently have you been called upon to defend your opinion on issues related to climate-change?”

Attack. “In the past, how frequently have you attempted to undermine your opponent's opinion on issues related to climate-change?” “In the past, how frequently have you attempted to confront your opponent's opinion on issues related to climate-change?”

Results

Predicting attitude polarization from advocacy style and initial stance. Similar to Study 1, attitude polarization was not predicted by advocacy style, $F(1, 174) = 1.37, p = .24$, initial stance, $F(1, 174) = 1.12, p = .29$, or the interaction between the two, $F(1, 174) = .62, p = .43$.

Principal components analysis. Similar to Study 1, an exploratory factor analysis again produced a two-component solution for meta-cognitive confidence. Component loadings from the pattern matrix indicated that yet again thought favorability (.80) and perceived source credibility (.63) loaded onto one component. However, this time, Component 1 contained confidence in positive thoughts (.28) which loaded equally strongly onto Component 2 (.27), along with perceived argument quality (.81), perceived effort (.75), thought conviction (.50), confidence in negative thoughts (.63).

Chapter 7 General Discussion

Introduction

Across seven studies and three separate research streams, we examine the mechanisms and consequences of pro-attitudinal advocacy on attitude change and communicative intentions. Given that people possess increasingly extreme attitudes, which leads to the formation of echo chambers, the driving force behind this thesis was to understand how we can depolarize attitudes and increase openness to the opposition. We do so by studying pro-attitudinal advocacy (advocating for one's position) in particular, because it is frequent in the current socio-political climate. Thus, our primary research question was: what factors and mechanisms in the context of pro-attitudinal advocacy lead to attitude depolarization and open communicative practices?

Given that highly convicted attitudes are difficult to change (Krosnick & Petty, 1995), we turn to alternative methods to depolarize attitudes and increase openness compared to standard persuasion, whereby a source communicates a persuasive message to a third-party recipient. The alternative that we turn to is self-persuasion, whereby the message-generating source and recipient are the same person. Thus, whether one is aware of it or not, the target of persuasion is the person generating the message itself. A plethora of self-persuasion paradigms have shown that attitude depolarization occurs as a result of: group discussion (group polarization; see Isenberg, 1986; Sunstein, 2002 for reviews), simply thinking about an attitude object (mere thought; see Tesser, 1978 for a review), generating reasons for an attitude (self-reflection; see Wilson, 1990 for a review), describing an attitude object (saying-is-believing; Higgins & Rholes, 1978), and generating persuasive messages for a discrepant position (counter-attitudinal advocacy; Hovland et al., 1953). However, we focus on one particular self-persuasion paradigm, in which people generate messages in favor of one's own

position, i.e. pro-attitudinal advocacy (Briñol et al., 2012), given that this paradigm is extremely understudied, but represents the most ecologically valid form of advocacy in today's society.

The Current Research

The two main goals of the current research were to: a) Understand the mechanisms which lead to depolarization and increased openness following a pro-attitudinal advocacy attempt, and b) Uncover how to re-frame pro-attitudinal advocacy attempts in ways which depolarize attitudes and encourage openness to the opposition. We examined these in the context of three separate research streams: (1) Meta-Cognitive Confidence and Self-Persuasion (Studies 1 and 2), (2) Moral versus Practical Advocacy (Studies 3-5), and (3) Attack versus Defense (Studies 6 and 7). In Stream 1, we sought to understand whether unconstrained pro-attitudinal advocacy indeed leads to depolarization, and why this may occur, by measuring meta-cognitive confidence in advocates' advocacy attempts.

Previous work supports a “saying-is-believing” effect, whereby people become more convinced, i.e. polarize, post-advocacy (see group polarization and saying-is-believing literature; Chapter 3). However, work in standard persuasion, and more recently self-persuasion, has shown that meta-cognitive confidence, or the amount of confidence people place in their arguments and thoughts, can lead people's attitudes to depolarize post-advocacy (Briñol et al., 2012; Clarkson et al., 2011, 2013). Thus, we examined the possibility that “saying-is-not-always-believing” in the context of pro-attitudinal advocacy.

In Study 1, we found evidence that attitudes depolarize if people place low confidence in their advocacy attempts. To our knowledge, this is the first study to demonstrate the effects of meta-cognitive confidence on depolarization in a pre-post study design, in which people advocate for their attitudes online, unconstrained. This finding is interesting in light of previous work, because most work in self-persuasion shows that people polarize when they

advocate for their opinion (see Chapter 3 for a review of self-persuasion), while – the current work shows the opposite. We find that people appear to question their attitudes, regardless of initial stance. Corroborating recent work in self-persuasion (see Briñol et al., 2012), we show that this depolarization is primarily driven by meta-cognitive confidence in one's advocacy attempt.

Further, we predicted that elaboration (amount of thinking) is a likely moderator of the effects of meta-cognitive confidence on attitudinal and behavioral outcomes, such that Need-for-Cognition (NfC) would moderate the relationships between meta-cognitive confidence and depolarization, and meta-cognitive confidence and intentions to engage with the opposition. This is because people who are high in NfC are more likely to use meta-cognitive confidence to inform their attitudes, and are also more open to revising their beliefs when confidence is low (Briñol & Petty, 2003; Tormala et al., 2002; Wegener & Petty, 1997). Results indicated that as predicted, NfC moderated the effects of meta-cognitive confidence on depolarization, such that, those who were high compared to low in NfC were more likely to depolarize post-advocacy if confidence in advocacy was low. Further, those who were high in NfC also reported greater willingness to engage with those holding opposing views, when confidence in advocacy was low. We show that NfC may importantly act as a buffer against polarization occurring post-advocacy, by translating low confidence into depolarization and greater willingness to interact with the opposition.

In Stream 2, we sought to understand the mechanisms and consequences of re-framing pro-attitudinal advocacy attempts in moral versus practical terms. This set of studies sought to explore ways in which advocacy may be re-framed in order to stimulate depolarization, increase openness or decrease closed mindedness post-advocacy. The distinction examined in Stream 2 was moral advocacy (grounded in moral values, ethics, virtues) versus practical advocacy (grounded in cost-benefit analyses and economic

consequences). This particular set of studies was motivated by our observation that advocacy crouched in moral rhetoric tends to yield polarized attitudes (Clifford, 2019; Clifford et al., 2015), and also increases people's willingness to proselytize their own opinions (Valenzuela et al., 2017). Thus, we sought to explore an alternative way of framing advocacy in order to encourage free speech while minimizing social conflict.

While previous work has acknowledged the destructive effects of moral rhetoric (see Brady et al., 2017; Kovacheff et al., 2018; Ryan, 2017), our set of studies was the first to: (1) Examine the effects of moral advocacy specifically in the context of self-persuasion, (2) Determine the specific mechanisms through which moral advocacy leads to negative consequences, and (3) Explore the positive effects of practical advocacy and its potential to bridge ideological divides. While previous work in standard persuasion has examined the consequences of moral reframing (Feinberg & Willer, 2013; 2015), and moral matching effects (matching messages to moral foundations, political orientation, or attitude bases; Luttrell et al., 2016; Luttrell et al., 2019), we were interested in a general advocacy framing regardless of these complexities, that might stimulate greater openness post-advocacy. We predicted that moral advocacy may drive polarization and proselytization due to one or more of the following mechanisms: (1) Reliance on the five moral foundations (collectively termed "moral expressiveness"), (2) Reliance on moral systems framing (deontology vs consequentialism), (3) Moral emotions, specifically those capturing moral outrage, and (4) Increased meta-cognitive confidence during advocacy.

Using text-based analysis, in Study 3 we found that those encouraged to engage in moral (compared to practical) advocacy wrote arguments capturing greater moral expressiveness, moral systems framing, and moral outrage towards an issue related to migration. In Study 4 we importantly found that only moral expressiveness was consequential in predicting attitude polarization. In Study 5 we replicate the superiority of moral

expressiveness over other moral content in explaining self-persuasion occurring during pro-attitudinal advocacy, by showing that moral expressiveness predicted proselytization intentions as well. While some effects of deontology and meta-cognitive confidence also emerged, these were not consistent across the studies. The results suggest moral expressiveness as the primary mechanism through which moral advocacy may lead to ideological divides, by increasing polarization and intentions to proselytize. More importantly, our results provide evidence for a novel advocacy paradigm, i.e. practical advocacy, which may attenuate the destructive effects of advocacy by decreasing polarization and proselytization.

Finally, in Stream 3 we sought to examine the consequences of pro-attitudinal advocacy beyond depolarization, by testing effects on behavioral intentions, such as, information-seeking preferences. We also sought to understand the consequences of re-framing pro-attitudinal advocacy attempts in ways which either attack the opposition (Attack) compared to justifying one's own position (Defense). Studies 6 and 7 represented our second attempt at manipulating or re-framing pro-attitudinal advocacy in ways which induce depolarization and open-up communicative intentions.

To do so, we distinguished between two possible types of pro-attitudinal advocacy: Attack versus Defense. In the Attack condition, advocates were asked to justify their position by degrading the opponent's position, and in the Defense condition advocates were asked to simply provide arguments for their own position. While we were not the first to make this distinction (see advocate vs refute work in the communications literature, and negative vs positive advertising in political advertising theory; Haynes & Rhine, 1998; Infante & Rancer, 1993; Kuhn, 1991), this work is the first to: (1) Define variations of pro-attitudinal advocacy (Attack vs Defense), compared to "free" or unconstrained advocacy in the field of social psychology, (2) Examine consequences of these variations in advocacy on speakers'

themselves (while previous work only examined antecedents of such advocacy, e.g. effects of trait-level argumentativeness; see Rancer & Avtgis, 2006 for a review).

Given that previous work suggests that: (1) Attack could lead to greater meta-cognitive confidence, polarization and proselytization than Defense due to people's difficulty generating arguments to support their own position (Wilson & Dunn, 1986; Limon et al., 2008), but that (2) Attack could also lead to increased confrontational emotions such as anger, which may paradoxically lead to uncongeniality bias or seeking attitude-inconsistent information (Young et al., 2011; Kim, 2016), our key question was: Can Attack promote greater or less openness than Defense, and if so, why? In Study 6 we found that Attack advocates were three times more likely to seek attitude-inconsistent information (over attitude-consistent information) compared to their Defense counterparts. Interestingly, this was not due to attitude polarization, meta-cognitive confidence or emotions such as, anger.

Thus, in Study 7, we sought to examine an alternative explanation for this finding. By measuring advocates' levels of defense motivation (motivation to justify one's position) and accuracy motivation (motivation to seek more accurate worldviews), we found our effect to be driven by increased accuracy motivation among Attack advocates. This is likely explained by the fact that Attack draws advocates' attention to the opponent's position and highlights the utility of novel information to be gained by paying attention to the opposition's views. Our findings suggest that encouraging advocates to tune-in to opposing views may encourage more positive information-seeking practices.

Theoretical Implications

Given the findings across all three research streams, a few conclusions can be drawn regarding pro-attitudinal advocacy, more generally. Specifically, three key findings emerge, which carry implications for the study of pro-attitudinal advocacy, and self-persuasion: (1) Self-generated persuasion following pro-attitudinal advocacy is driven by a range of

mechanisms, including motivations and message content, (2) Pro-attitudinal advocacy can directly impact communicative intentions, without changing attitudes, and (3) Pro-attitudinal advocacy can be re-framed in ways which decrease polarization and increase open communicative intentions. We will examine each of these claims in-turn.

Multiple Mechanisms Underlying Self-Persuasion. While the limited research studying pro-attitudinal advocacy highlights meta-cognitive confidence as a mechanism impacting attitude depolarization (see Briñol et al., 2012), we find evidence for a range of factors influencing attitude depolarization, such as, moral expressiveness. Previous work shows that low meta-cognitive confidence, as indexed by perceived effort and perceived argument quality, predicts depolarization post-advocacy (Briñol et al., 2012). This is an important finding which corroborates work in standard persuasion, demonstrating the effects of meta-cognitive confidence on attitude change (see Petty et al., 2007 for a review), and also work in self-persuasion (e.g. mere thought paradigm; Clarkson et al., 2011, 2013; Tesser, 1978), demonstrating the effects of meta-cognitions on attitude depolarization.

The current research corroborates these findings by showing that meta-cognitive confidence (as indexed by a composite of various meta-cognitions, including perceived effort, perceived argument quality, thought confidence, and perceived knowledgeability) predicts attitude depolarization following pro-attitudinal advocacy (see Stream 1, Studies 1 and 2). Specifically, we show that low meta-cognitive confidence predicts attitude depolarization post-advocacy, and that individual differences such as NfC, moderate such effects, such that those high in NfC are more likely to depolarize and engage with dissimilar others when meta-cognitive confidence is low.

What this means is that level of elaboration, or an advocate's propensity to engage in effortful deliberation (see Elaboration Likelihood Model; Petty & Briñol, 2014; Petty & Briñol, 2012) is a possible mechanism which may account for the effects of pro-attitudinal

advocacy on self-persuasion outcomes. Given that NfC is used as a proxy for depth of processing, our findings suggest that confidence-depolarization effects may partially depend upon elaboration. Dual-process theories highlighted by the standard persuasion literature (see Chen & Chaiken, 1999; Petty & Briñol, 2014 for reviews) show that attitude change may occur via two “routes” to persuasion depending on the recipient’s level of elaboration. When level of elaboration is high (e.g. high NfC) recipients tend to follow a central or systematic route, whereby attitude change is influenced by thoughtful processes, such as, processing the quality of arguments presented. When level of elaboration is low (e.g. low NfC), recipients tend to follow a peripheral or heuristic route, whereby attitude change is influenced by superficial cues, such as, mere number of arguments contained in the message, regardless of their merit. Drawing on these dual routes to persuasion, our findings imply that attitude change occurring via self-persuasion may rely more on controlled rather than automatic processes. That is, confidence-depolarization effects and confidence-behavioral intention links may not be automatic, but rather, depend upon thoughtful deliberation. Future research manipulating confidence is required to validate this claim.

In addition to meta-cognitive confidence, other mechanisms may also drive self-persuasion effects occurring during pro-attitudinal advocacy. While previous work emphasizes the effects of meta-cognitive confidence (see Gordijn et al., 2011 for an exception, showing perceived threat effects), we do not know what other mechanisms may drive depolarization and communicative intentions in the context of pro-attitudinal advocacy. Importantly, we find evidence that self-persuasive effects of advocacy may be driven by lesser-known mechanisms, such as, message content.

For example, we find that moral expressiveness (tendency to express words related to the five moral foundations), accounts for the effects of moral versus practical advocacy on polarization. Compared to those who engaged in moral advocacy, advocates who engaged in

practical advocacy were more likely to depolarize post-advocacy, driven by lower moral expressiveness in their arguments, compared to other types of moral content (e.g. moral emotions). These studies collectively suggest that mechanisms other than confidence-related variables, such as, moral expressiveness, likely predict self-persuasion outcomes following pro-attitudinal advocacy.

It is important to note that the current research examined the impact of a *single* pro-attitudinal advocacy attempt on self-persuasion, *immediately* following argument generation. There is outstanding work to be done to assess the impact of a) repeated advocacy attempts, b) time passed following an advocacy attempt on attitude change, and c) longevity of attitude change (i.e. how long attitude change lasts for and if there are any rebound effects). Whether mechanisms such as meta-cognitive confidence and moral expressiveness continue to impact attitude change in the afore-mentioned circumstances remains to be investigated as well.

Direct Impact on Communicative Intentions. Although a range of factors may directly influence attitude change, it is not necessary to produce changes in communicative intentions in the context of pro-attitudinal advocacy. While (de)polarization was a key dependent variable examined across most studies, we also examined communicative intentions, such as, willingness to engage with the opposition (Stream 1), intentions to proselytize one's opinions (Stream 2), and information-seeking preferences (Stream 3). Although depolarization occurs in many of our studies (Stream 1, Studies 1 and 2; Stream 2, Study 4), it also does not occur in others (Stream 2, Study 5; Stream 3, Studies 6 and 7). Although depolarization does not occur in some of our studies, we still observe changes in communicative intentions, suggesting that pro-attitudinal advocacy may have a direct impact on behavioral intentions, without changing attitudes.

What this means is that depolarization may not be a pre-requisite for behavioral intentions and behaviors to change. Social psychological theories of attitude and behavior

change posit that behaviors may change not only as a result of attitude change, but also due to changes in perceived norms, and efficacy beliefs, for example (see Theory of Planned Behavior; Ajzen, 1991; Ajzen & Fishbein, 2005). In the context of pro-attitudinal advocacy, it is possible that attitudes need not shift in order to observe changes in communicative intentions. For example, we show that proselytization intentions are influenced by moral advocacy, even without attitudes polarizing post-advocacy (see Stream 2, Study 5), suggesting that communicative intentions may change as a result of phenomena other than for attitudes change.

We also show that information-seeking preferences and receptivity to opposing views is directly influenced by Attack vs Defense advocacy styles, even without attitudes changing (see Stream 3, Studies 6 and 7). This behavioral change in absence of attitude change corroborates recent work suggesting that behaviors towards traditionally prejudiced groups (e.g. gay people) may change due to perceived norm changes (i.e. passing of the same-sex marriage legislation), even if prejudiced attitudes towards these people do not change (Tankard & Paluck, 2017). In the context of pro-attitudinal advocacy and self-persuasion, we find that attitude change is not a pre-requisite for behavioral intention change, suggesting that behaviors themselves may also change post-advocacy without attitudes changing.

One possible mechanism other than for attitudes, through which behavioral intentions and behaviors may change, is meta-cognitive confidence. While meta-cognitive confidence did not emerge as the key mediator of behavioral intentions across all research streams, it did, however, by-pass attitude change in predicting behavioral intentions in some capacity. In Stream 1, meta-cognitive confidence predicted willingness to engage with the opposition directly, and not via attitude depolarization. In Stream 2, meta-cognitive confidence predicted intentions to proselytize one's opinions, even when attitudes remained unchanged overall. These findings suggest that lack of confidence might make one more wary of pushing one's attitudes on others, even if it does not weaken the attitude itself. Further, we show that meta-

cognitive confidence when interacting with other individual dispositions, such as, level of elaboration, may manifest as approach-motivated actions, such as willingness to engage with the opposition.

This corroborates previous work on attitude certainty and pro-attitudinal advocacy, which shows that advocates placing high certainty in their attitudes tend to engage in greater proselytization compared to those who possess low-moderate attitude certainty (Cheatham & Tormala, 2015; 2017). Our work also corroborates research on consumer self-confidence and decision-making, which demonstrates that consumers with high self-confidence possess greater confidence in seeking and integrating relevant consumer information (Clark, Goldsmith, & Goldsmith, 2008), and also display greater information-seeking tendencies (Loibl, Cho, Diekmann, & Batte, 2009). More importantly, this work suggests that consumers with high self-confidence are more likely to share their knowledge with others, and also more likely to influence other consumers' purchase behavior compared to those who are low in self-confidence (Bearden, Hardesty, & Rose, 2001).

In addition to the effects of confidence, another mechanism which may drive behavioral intentions, is motivations. Specifically, we find that accuracy motivation (motivation to hold accurate views) compared to defense motivation (motivation to confirm existing views; see Hart et al., 2009 for a review) accounts for the effects of Attack vs Defense advocacy styles on information-seeking preferences and receptivity to opposing views, even if attitudes do not change (see Stream 3, Studies 6 and 7). Attack advocates were more likely to seek attitude-disconfirming information, select partners holding opposing views, and express greater receptivity to opposing views and express positive perceptions of the opposition, driven by increased accuracy motivation post-advocacy, compared to Defense advocates.

Re-framing Pro-attitudinal Advocacy Attempts. According to previous work in self-persuasion, unconstrained advocacy, during which people freely generate arguments for

their own position, typically leads to polarization (see Chapter 3 for a review). However, work in pro-attitudinal advocacy, specifically, shows that compared to “free” or unconstrained advocacy, re-framing advocacy in particular ways “nudges” people towards depolarization (see Briñol et al., 2012; Gordijn et al., 2001). Similarly, the current research suggests that encouraging advocates to generate arguments in specific ways may lead to depolarization and also open communicative intentions.

Reconciling opposing views is extremely cognitively demanding (Kuhn, 2019), and unfortunately may lead to back-fire effects in perspective-taking tasks, such as in counter-attitudinal advocacy paradigms (Catapano et al., 2019). While a meta-analysis of counter-attitudinal advocacy suggests that dissonance-arousing processes generally lead to attitude depolarization (Kim et al., 2014), more recent work suggests that counter-attitudinal advocacy may indeed back-fire and reduce receptiveness to opposing viewpoints, as value incongruence is highlighted in the process of deliberating the opposing view (Catapano et al., 2019). Interestingly, modifying the advocacy context such that advocates have the opportunity to briefly affiliate with someone holding the opposing viewpoint, but similar overall values (e.g. political values) reverses this effect and induces depolarization post-counter-attitudinal advocacy.

Thus, it appears that specific modifications to the advocacy context may lead to attitude depolarization and more open communicative intentions. For example, recent work in group polarization shows that introducing a moderator to group discussions stimulates greater depolarization than discussion without a moderator (Strandberg et al., 2017). Similarly, previous work in audience design shows that the “saying-is-believing” effect is attenuated when advocates are asked to generate message descriptions for three-person as opposed to single-person audiences, unless audience validation is provided (Hausmann et al., 2008). Although work in pro-attitudinal advocacy does not demonstrate variations in self-persuasion depending on variations in target audience (e.g. family, friend, professor; see Briñol et al.,

2012), other modifications to the advocacy context have been successful.

For example, Briñol and colleagues (2012) show that asking advocates to target their advocacy attempt at themselves as opposed to another person leads to greater depolarization, via lower perceived effort expended in argument generation. Similarly, previous work undertaken by Gordijn and colleagues (2011) shows that informing advocates that their attitude reflects the minority position compared to the majority position leads to depolarization, possibly due to concerns about social validation. Other than for these two manipulations, to our knowledge, no other work has examined the consequences of re-framing advocacy appeals in ways which promote depolarization and open communicative intentions, in the context of pro-attitudinal advocacy.

The current research provides evidence for two novel advocacy paradigms which may be used to encourage advocacy in more positive ways. While we do not find variations in target audience (i.e. whether an audience agrees, disagrees, or is neutral towards the issue) to influence attitude depolarization (Stream 1 studies, Chapter 4, see Supplementary Materials), we find that moral vs practical advocacy (see Stream 2 studies), and Attack vs Defense (see Stream 3 studies) influences attitude depolarization and communicative intentions. Specifically, we find that asking advocates to generate practical compared to moral reasons leads to lower polarization and proselytization intentions post-advocacy. We also find that asking advocates to focus on their opposition's faults, compared to justifying one's own position, leads to uncongeniality bias and increased receptivity to opposing views.

Practical Implications

In addition to theoretical implications, this research carries practical implications for how pro-attitudinal advocacy can be leveraged to bridge ideological divides. Given the increasing prevalence of online advocacy (Hestres, 2015) on social media sites, such as, Reddit and Facebook, we pose the following question: can we moderate advocacy in ways which encourage free speech whilst simultaneously minimizing conflict between opposing

groups? A moderated Reddit forum, “Ask Trump Supporters” comprising discussion between Republicans and Democrats in the wake of the 2016 Presidential election provides some support for this notion (Reddit, 2016). Rules for conversation guide the forum, such that advocates are not allowed to use profanity, or directly condemn others in the forum, but rather, clarify their political arguments by posing questions to dissimilar others. While to our knowledge no formal research directly examines the consequences of this forum on communication practices, our observation of conversations in the forum indicate more peaceful and collaborative discourse than otherwise expected from politically charged advocates.

The findings of the current research may be relevant to policymakers, NGO’s, and educational institutions, for example, that have a key interest in encouraging civil political discourse. While the current work did not directly examine naturalistic online interactions such as on Reddit, there are a few key manipulations to the advocacy context which can be made based on our findings, which may induce depolarization or increase openness to the opposition.

Induce Uncertainty in People’s Arguments. One way of inducing depolarization is by subtly leading people to question their arguments by lowering meta-cognitive confidence in their advocacy attempts. This recommendation is partially supported by Stream 1 Study 1 and Study 2, in which we demonstrated that low meta-cognitive confidence predicts attitude depolarization. Although we did not directly manipulate meta-cognitive confidence, our findings suggest that inducing low meta-cognitive confidence may lead to depolarization post-advocacy. Previous work in standard persuasion has shown that decreasing meta-cognitive confidence (e.g. thought confidence) using false feedback tends to lead to attitude depolarization (Petty, et al., 2002; Tormala et al., 2004). That is, providing advocates with feedback that their arguments or thoughts are “weaker” or “different” compared to others, likely induces low meta-cognitive confidence, and thus depolarization. This is similar to

numerical support effects, in which Gordijn and colleagues (2011) provided feedback on whether an advocate's position was the majority or minority position (with the latter information leading to higher rates of depolarization). Whilst false feedback may be deceptive (e.g. if the feedback provided is untrue) and thus raise ethical concerns, it may be possible to use false feedback for attitude issues which have empirical support (e.g. providing accurate information on the increasing support for gay marriage in the US successfully shifted norm perceptions; Tankard & Paluck, 2017).

In our studies, the manipulations which produced the largest effects on meta-cognitive confidence were the “free” or unconstrained advocacy manipulation, and the advocacy manipulation aimed at different target audiences (see Stream 1, Chapter 4; note, no variation in confidence based on variations in target, but overall effects on confidence). These manipulations may be leveraged in ways which promote attitude depolarization, if followed by a false feedback task post-advocacy. Providing feedback on the quality of arguments or content of thoughts is likely to have the greatest effect on attitude depolarization, given that perceived argument quality and thought confidence were the only two specific meta-cognitions which uniquely correlated with depolarization consistently across both Studies 1 and 2 ($.16 < r_{\text{argument quality}} < .17$, $.11 < r_{\text{thought confidence}} < .12$).

Request Arguments to be Framed in Practical Terms. Another method of increasing depolarization and reducing closed communicative practices, such as, proselytization, is by encouraging advocates to generate advocacy appeals grounded in practical as opposed to moral reasons. This recommendation is supported by Stream 2 Studies 3, 4, and 5, in which we demonstrated that practical advocacy leads to lower polarization and decreased intentions to proselytize, compared to moral advocacy, due to lower expression of moral values. While previous work suggests that non-moral rhetoric may reduce polarization (Kovacheff et al., 2018), the current research is the first to show, that in the context of advocacy, re-framing persuasive appeals in practical terms (i.e. based on

economic consequences, cost-benefit analyses, pragmatism), compared to moral terms leads to less polarization and proselytization. Given that people often rely on moral values to justify their attitudes (Clifford et al., 2015; Tappin & McKay, 2019), and perceive their own moral values to be universal and sacred (Tetlock, 2003), encouraging advocacy appeals to be grounded in reasons other than for moral values leads to less negative outcomes post-advocacy, as advocates may not be accustomed to deliberating non-moral, pragmatic reasons for their attitudes. The benefit of encouraging practical advocacy is that it allows people to retain their right to free speech, but simultaneously decreases the likelihood of negative behaviors which lead to social division.

Encourage Focus on Opponent's Arguments. A final method of increasing open communicative practices, is by encouraging advocates to pick apart their opponent's arguments, as opposed to providing arguments for their own position. This recommendation is supported by Stream 3 Studies 6 and 7, in which we demonstrated that Attack advocates were more prone to uncongeniality bias, as indexed by their preference for attitude-inconsistent information, willingness to interact with dissimilar others, and expression of a more open mindset compared to Defense advocates. While passive exposure to opposing views on social media can lead to polarization (Bail et al., 2018), we suggest that actively deliberating the opposition's position may lead to more open communicative practices. This is likely because advocates do not frequently contemplate the reasons for an opponent's position, and thus may find it difficult to attack those specific reasons.

Our findings demonstrating uncongeniality bias is important, because congeniality bias has been shown to lead to polarized attitudes, while discussion with dissimilar others weakens this link (Kim, 2015). Then, it follows that if advocacy attempts themselves encourage advocates to interact with dissimilar others, downstream effects on selective exposure and polarization, for example, can be attenuated. People often underestimate the

polarizing effects of justifying their own position, especially in a group (Keating, Van Boven, & Judd, 2016), thus we provide an alternative form of advocacy which may lead to more positive outcomes post-advocacy.

Limitations and Future Directions

The current research provides novel insights on the mechanisms underlying pro-attitudinal advocacy, and how advocacy can be re-framed in ways which reduce negative outcomes, such as, attitude polarization, congeniality bias, and proselytization. However, a few limitations must be acknowledged. We will acknowledge limitations and provide future directions within five key categories: (1) Replicability of depolarization effects, (2) Correlation vs causation, (3) Generalizability to in-person interactions, (4) Distinction between behaviors versus behavioral intentions, and (5) Using text-based analysis in self-persuasion paradigms.

Replicability of Depolarization Effects. Although attitude depolarization was demonstrated across some studies (e.g. in Studies 1, 2, and 4), this did not occur consistently across all studies. There are a few possible reasons for this observation. First, it is possible that not all advocacy manipulations are equally able to induce depolarization. While free advocacy in Stream 1 was able to induce depolarization, Attack vs Defense advocacy styles did not induce depolarization, but rather, directly impacted uncongeniality bias. This is consistent with early work in counter-attitudinal advocacy, showing different rates of depolarization using improvised versus non-improvised advocacy (Janis & King, 1954; King & Janis, 1956), work in standard persuasion showing different attitude change results depending on the type of moral values portrayed in advocacy appeals (Feinberg & Willer, 2013; 2015), and work in political psychology demonstrating differences in attitude change depending on whether the messages presented contained moral or practical arguments (Leidner et al., 2018). It is possible that not all advocacy appeals equally impact attitude

depolarization. Nevertheless, if advocacy appeals are able to stimulate behavioral intention-change at the least, as in the current research, this is promising.

Second, it is possible that particular sample sizes were insufficient to detect links between our variables and attitude depolarization. This is consistent with previous work in self-persuasion, demonstrating that most people's attitudes do not typically depolarize following a self-persuasion intervention (approximately 15% depolarization rate; Tesser, 1978). In Stream 2, Study 5 (see Chapter 5 on moral advocacy and polarization), although we did not find a correlation between moral expressiveness and attitude depolarization in the single study itself, a mini meta-analysis across both studies indeed revealed an overall significant effect. Thus, it is possible that some sample sizes were sufficient to detect between-condition effects in constructs such as moral expressiveness, and variation in behaviors such as information-seeking preferences, but were not sufficient to detect variation in depolarization.

Third, it is possible that depolarization is less likely to occur in certain populations compared to others, and also certain attitude issues compared to others. In the current research, depolarization rates (percentage of people depolarizing) were higher in studies which employed university samples (range: 30% - 35%) compared to those which employed MTurk samples (range: 10% - 33%). It is possible that age influences the extent to which attitudes are malleable, such that younger individuals are more open to attitude change (Krosnick & Alwin, 1989), although we did not find a significant correlation between age and depolarization in any of our studies ($-.09 < r < .08$.) This is consistent with work supporting the "lifelong openness model" of attitude change, which demonstrates that attitudes of older individuals change as much as younger individuals, when personal experiences or encounters with the attitude object are taken into account (Tyler & Schuller, 1991).

It is also possible that political orientation may influence attitude depolarization (Feinberg & Willer, 2013, 2015), however, we did not find political orientation to moderate our effects, as shown by the results of Stream 2 (moral vs practical advocacy studies). We are unable to pinpoint the source of variation in attitude depolarization given that: (1) Age and political orientation are unlikely to be moderators of our effects (in Stream 2 research we formally tested effects of political orientation and did not find it to moderate effects of advocacy on polarization and proselytization), and that (2) A meta-analysis across the studies would be less interpretable due to substantial differences in study design (e.g. free advocacy vs moral/practical vs Attack/Defend), and that (3) Attitude depolarization did not consistently occur for each attitude issue tested (e.g. for the climate-change issue depolarization occurred in Studies 1, 2, and 4, but not in studies 6 and 7). Future studies may examine other moderators of attitude depolarization, such as sample demographics, and type of attitude issue tested, given that issues tend to vary in their level of moral conviction (e.g. migration issues tend to be more morally ingrained compared to climate-change issues; see Stream 2 studies), and their ability to evoke emotional reactions (Reeves et al., 2016).

Correlation vs Causation. While the current work establishes correlational evidence for the effects of meta-cognitive confidence and message content on attitude depolarization for example, we are unable to make causal claims about the effectiveness of these methods in inducing depolarization. This is an obvious limitation of using cross-sectional research to test causal process (see Kline, 2015). However, the current work is an important step in identifying particular variables (e.g. meta-cognitive confidence, moral message content) which may predict depolarization and communicative intentions such as proselytization, specifically in the context of pro-attitudinal advocacy. Future research is required to manipulate such variables and others, in order to establish causal relationships.

Construct Distinctiveness. Although we found relationships between meta-cognitive confidence and attitude depolarization in the current research, no single meta-cognitive confidence measure was solely responsible for this effect. Rather, it was a combination of different meta-cognitions associated with each other which predicted depolarization. Previous work tends to examine the various meta-cognitive variables in isolation (e.g. perceived argument quality and perceived effort; see Briñol et al., 2012; thought confidence, see Tormala et al., 2002, Tormala et al., 2007), while the current research tested the effects of multiple variables. While past work also demonstrates collinearity between the various meta-cognitive variables (e.g. perceived argument quality and perceived effort; see Briñol et al., 2012), it may be worthwhile to test for construct distinctiveness and the unique predictive value of each of these variables.

Generalizability to In-Person Conversations. Another limitation of the current work is that all studies were conducted online, as opposed to in-person, thus the generalizability of our effects is questionable. Previous work in the computer-mediated communication literature suggests that online versus in-person interactions have different outcomes on attitudes and behaviors (Ho & McLeod, 2008; Lea & Spears, 1991; Spears, Lea, & Lee, 1990). For example, online discussion tends to yield more polarized attitudes compared to in-person discussion, possibly due to lack of social cues (Lea & Spears, 1991). Given these differences, it is possible that in-person advocacy may have different consequences compared to online advocacy. In-person interactions are also more complex and difficult to study, given the contextual cues that have to be taken into account (e.g. appearance and demographic of the target of advocacy, tone, voice, and pace of speech). Thus, for a preliminary investigation of pro-attitudinal advocacy, online conditions provide a more suitable experimental platform, free from such confounding variables which may attenuate or exacerbate our effects. Further, online studies allow us to mimic the increasing

advocacy which happens on social media platforms such as Reddit, Twitter, and Facebook.

Future research may examine variations in the mechanisms and consequences of pro-attitudinal advocacy in online versus face-to-face contexts.

Behaviors vs Behavioral Intentions. Another limitation is that the current research primarily investigated behavioral intentions as opposed to actual behaviors. With the exception of information-seeking preferences in Stream 3 (Studies 6 and 7), most dependent variables examined behavioral intentions, such as intentions to proselytize opinions, and willingness to interact with dissimilar others. Importantly, the current work endeavored to capture intentional behavior, and not habitual behavior (see Ouellette & Wood, 1998 for a review of conscious versus automatic behavior) in which advocates make premeditated or conscious attempts to seek particular kinds of information, or interact with particular kinds of people, for example. While the strongest predictor of intentional behaviors is behavioral intentions (see Ajzen, 1991; Ajzen & Fishbein, 2005), future research should examine actual behaviors post-advocacy, such as signing petitions, and voting behavior (see Allpress, Barlow, Brown, & Louis, 2010; Cheatham & Tormala, 2015, 2017; Skitka & Mullen, 2002).

Using text-based analysis. A final recommendation for future studies is the exploration of text-based analysis to examine the consequences and mechanisms underlying pro-attitudinal advocacy. Typical work in advocacy simply examines the effects of advocacy on polarization, for example, but text-based analysis may provide insights which we are unable to gain using existing measures. For instance, in Stream 2 (moral vs practical advocacy), the non-text-based measures, such as meta-cognitive confidence typically employed in advocacy research, did not emerge as consistent predictors of polarization and proselytization across the studies. However, the text-based measure capturing moral expressiveness emerged as a consistent predictor of our effects. Thus, text-based analysis

may be a useful complement to existing measures and add more nuance to self-persuasion studies in the future.

Concluding Remarks

The current work furthers our understanding of self-persuasion occurring during pro-attitudinal advocacy and sheds light on the mechanisms and consequences of advocacy on practices relevant to bridging ideological divides. This research importantly identifies ways in which practitioners may encourage free speech in ways that minimize social conflict. Thus, we propose pro-attitudinal advocacy and other related self-persuasion paradigms as potentially effective complements to standard persuasion methods to encourage attitude and behavior change. The possibility of harnessing self-persuasion to bridge polarized groups is undoubtedly intriguing in the current political climate.

Note, the implication throughout this work is that depolarization and accompanying open communicative behaviors are desirable due to their ability to bring opposing groups together, decrease dogmatism and increase collaborative action. However, there may be circumstances under which polarization and proselytization are more desirable instead (e.g. health behavior contexts). The current research focused on minimizing conflict in political contexts, thus the *value* of depolarization (a largely normative discussion) was out of scope. However, exploring the boundary conditions of the value of depolarization is an interesting future direction for both psychology and philosophy research.

Another interesting and important extension of the current research is examining the dyadic and group-level consequences of pro-attitudinal advocacy. That is, understanding how micro-level processes may feed into macro-level consequences for both advocates and their potential audiences. While self-persuasion is occurring in an advocate, audience attitudes may be shaped as well – this could result in a double polarization or depolarization effect or result in an asymmetrical pattern of attitude change. This dyadic-level impact of advocacy is out of scope for the current research which focuses

exclusively on self-generated persuasion. However, there is a deeper question to be explored on dyadic polarization and depolarization, and how self-persuasion versus standard persuasion may feed into this process. Thus a fruitful exploration of future research may include examining the effects of embedding self-persuasion in dyadic and social network contexts.

Chapter 8 References

- Abeywickrama, R. S., & Laham, S. M. (2017). Thought conviction as a meta-cognitive moderator of self-persuasion in counter-attitudinal advocacy. Unpublished Data.
- Abeywickrama, R. S., & Laham, S. M. (2020a). Meta-Cognition Predicts Attitude Depolarization and Intentions to Engage with the Opposition Following Pro-Attitudinal Advocacy. *Social Psychology*. <https://doi.org/https://doi.org/10.1027/1864-9335/a000424>
- Abeywickrama, R. S., & Laham, S. M. (2020b). Attack versus Defense. *Basic and Applied Social Psychology* (in progress).
- Abeywickrama, R. S., Laham, S. M., & Crone, D. L. (2018). Immigration and receiving communities: The utility of threats and emotions in predicting action tendencies toward refugees, asylum-seekers and economic migrants. *Journal of Social Issues*, 74(4), 756-773. <https://doi.org/10.1111/josi.12297>
- Abeywickrama, R. S., Rhee, J. J., Crone, D. L., & Laham, S. M. (2020). Why Moral Advocacy Leads to Polarization and Proselytization: The Role of Self-Persuasion. *Journal of Social and Political Psychology*, 8(2), 473–503. <https://doi.org/https://doi.org/10.5964/jspp.v8i2.1346>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of attitudes* (pp. 173–221). Mahwah, NJ: Lawrence Erlbaum Associates.
- Alba, J. W., & Marmorstein, H. (1987). The Effects of Frequency Knowledge on Consumer of Decision Making. *Journal of Consumer Research*, 14(1), 14–25. Retrieved from <http://www.jstor.org/stable/2489239>

- Albarracin, D., Johnson, B. T., & Zanna, M. P. (2014). *The Handbook of Attitudes* (pp. 35-90). New York: Psychology Press.
- Allport, G. W. (1935). Attitudes. In *A Handbook of Social Psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- Allpress, J. A., Barlow, F. K., Brown, R., & Louis, W. R. (2010). Atoning for Colonial Injustices: Group-Based Shame and Guilt Motivate Support for Reparation. *International Journal of Conflict and Violence*, 4(1), 75–88. <https://doi.org/10.4119/ijcv-2816>
- Anderson, S. E., Potoski, M., DeGolia, A., Gromet, D., Sherman, D., & Van Boven, L. (2014). Mobilization, Polarization, and Compromise: The Effect of Political Moralizing on Climate-change Politics. *APSA 2014 Annual Meeting Paper*.
- Araida, S., Chiacchiaro, E., Garney, G., Lievens, E., Ramanujam, S., & Sinclair, L. (2016). *Electoral Violence: A Study of the 2016 United States Presidential Election*. Retrieved from https://government.georgetown.edu/sites/government/files/files/upload/georgetown_-_2016_-_electoral_violence_group_project_1_1.pdf
- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. *Advances in Experimental Social Psychology Volume 4*(4), 1–34. [https://doi.org/10.1016/S0065-2601\(08\)60075-1](https://doi.org/10.1016/S0065-2601(08)60075-1)
- Axson, D., Yates, S., & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *Journal of Personality and Social Psychology*, 53(1), 30–40. <https://doi.org/10.1037/0022-3514.53.1.30>
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., ... Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences of the United States of America*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>

- Bearden, W. O., Hardesty, D. M., & Rose, R. L. (2001). Consumer Self-Confidence: Refinements in Conceptualization and Measurement. *Journal of Consumer Research*, 28(1), 121–134. <https://doi.org/10.1086/321951>
- Bebermeier, S., Echterhoff, G., Böhner, G., & Hellmann, J. H. (2015). The generalization of shared reality: When communication about one target shapes evaluations of other targets. *European Journal of Social Psychology*, 45, 623–640. <https://doi.org/10.1002/ejsp.2115>
- Bell, A. (1984). Language Style as Audience Design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/S004740450001037X>
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 183–200. <https://doi.org/10.1037/h0025146>
- Berger, J. (2011). Arousal Increases Social Transmission of Information. *Psychological Science*, 22(7), 891–893. <https://doi.org/10.1177/0956797611413294>
- Bizer, G. Y., Larsen, J. T., & Petty, R. E. (2011). Exploring the valence-framing effect: Negative framing enhances attitude strength. *Political Psychology*, 32(1), 59–80. <https://doi.org/10.1111/j.1467-9221.2010.00795.x>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., Van Bavel, J. J., & Fiske, S. T. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618923114>
- Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, 49(5), 811–821. <https://doi.org/10.1016/j.jesp.2013.04.005>
- Brandt, M. J. (2017). Predicting Ideological Prejudice. *Psychological Science*, 28(6), 713–722. <https://doi.org/10.1177/0956797617693004>

- Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The Ideological-Conflict Hypothesis: Intolerance Among Both Liberals and Conservatives. *Current Directions in Psychological Science*, 23(1), 27–34.
<https://doi.org/10.1177/0963721413510932>
- Brehm, S., & Brehm, J. (1981). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.
- Briñol, P., McCaslin, M. J., & Petty, R. E. (2012). Self-generated persuasion: Effects of the target and direction of arguments. *Journal of Personality and Social Psychology*, 102(5), 925–940. <https://doi.org/10.1037/a0027231>
- Briñol, P., & Petty, R. E. (2003). Overt head movements and persuasion: a self-validation analysis. *Journal of Personality and Social Psychology*, 84(6), 1123–1139.
<https://doi.org/10.1037/0022-3514.84.6.1123>
- Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, 20(5), 49–96.
<https://doi.org/10.1080/10463280802643640>
- Briñol, P., & Petty, R. E. (2015). Elaboration and validation processes: Implications for media attitude change. *Media Psychology*, 18(3), 267–291.
<https://doi.org/10.1080/15213269.2015.1008103>
- Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness versus sadness as a determinant of thought confidence in persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 93(5), 711–727. <https://doi.org/10.1037/0022-3514.93.5.711>
- Briñol, P., Petty, R. E., & Requero, B. (2017). Aggressive primes can increase reliance on positive and negative thoughts affecting self-attitudes. *Self and Identity*, 16(2), 194–214.
<https://doi.org/10.1080/15298868.2016.1255251>

- Briñol, P., Petty, R. E., & Tormala, Z. L. (2004). Self-Validation of Cognitive Responses to Advertisements. *Journal of Consumer Research*, 30(4), 559–573.
<https://doi.org/10.1086/380289>
- Burnstein, E., & Vinokur, A. (1973). Testing two classes of theories about group induced shifts in individual choice. *Journal of Experimental Social Psychology*, 9(2), 123–137.
[https://doi.org/10.1016/0022-1031\(73\)90004-8](https://doi.org/10.1016/0022-1031(73)90004-8)
- Burnstein, E., & Vinokur, A. (1975). What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology*, 11(5), 412–426.
[https://doi.org/10.1016/0022-1031\(75\)90045-1](https://doi.org/10.1016/0022-1031(75)90045-1)
- Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, 13(4), 315–332. [https://doi.org/10.1016/0022-1031\(77\)90002-6](https://doi.org/10.1016/0022-1031(77)90002-6)
- Cacioppo, J. T., & Petty R.E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology*, 42(a), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cacioppo, J. T., & Petty, R. E. (1989). Effects of Message Repetition on Argument Processing, Recall, and Persuasion. *Basic and Applied Social Psychology*, 10(1), 3–12.
<https://doi.org/http://dx.doi.org/10.1037/0022-3514.37.1.97>
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253. <https://doi.org/10.1037/0033-2909.119.2.197>
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306-307.
https://doi.org/10.1207/s15327752jpa4803_13

- Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), 1032–1043. <https://doi.org/10.1037/0022-3514.51.5.1032>
- Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45(4), 805–818. <https://doi.org/10.1037/0022-3514.45.4.805>
- Carver, C. S., & Harmon-Jones, E. (2009). Anger Is an Approach-Related Affect: Evidence and Implications. *Psychological Bulletin*, 135(2), 183–204. <https://doi.org/10.1037/a0013965>
- Catapano, R., Tormala, Z. L., & Rucker, D. D. (2019). Perspective Taking and Self-Persuasion: Why “Putting Yourself in Their Shoes” Reduces Openness to Attitude Change. *Psychological Science*, 30(3), 424–435. <https://doi.org/https://doi.org/10.1177/0956797618822697>
- Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J.M. Olson, & C. P. Herman (Eds.), *Social influence: the Ontario Symposium* (Vol 5, pp. 3–39).
- Chaiken, S., & Eagly, A. H. (1983). Communication modality as a determinant of persuasion: The role of communicator salience. *Journal of Personality and Social Psychology*, 45(2), 241–256.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460–473. <https://doi.org/10.1037/0022-3514.66.3.460>

- Chaiken, S., & Stangor, C. (1987). Attitudes and attitude change. *Annual Review of Psychology*, 38(1), 575–630.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. Uleman & J. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.
- Cheatham, L., & Tormala, Z. L. (2015). Attitude Certainty and Attitudinal Advocacy: The Unique Roles of Clarity and Correctness. *Personality and Social Psychology Bulletin*, 41(11), 1537–1550. <https://doi.org/10.1177/0146167215601406>
- Cheatham, L., & Tormala, Z. L. (2017). The Curvilinear Relationship Between Attitude Certainty and Attitudinal Advocacy. *Personality and Social Psychology Bulletin*, 43(1), 3–16. <https://doi.org/10.1177/0146167216673349>
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 73–96). New York: Guilford Press. <https://doi.org/10.4319/lo.2013.58.2.0489>
- Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated Heuristic and Systematic Processing. *Psychological Inquiry*, 10(1), 44–49. https://doi.org/10.1207/s15327965pli1001_6
- Cialdini, R. B., Petty, R. E., & Cacioppo, J. T. (1981). Attitude & Attitude Change. *Annual Review of Psychology*, 32, 357–404. <https://doi.org/doi:10.1146/annurev.ps.32.020181.002041>
- Clark, H. H., & Carlson, T. B. (1982). Hearers and Speech Acts. *Language*, 58(2), 332–373. <https://doi.org/10.1353/lan.1982.0042>

- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the Given-New Contract. *Discourse Production and Comprehension*, 1, 1–40. Retrieved from http://web.stanford.edu/~clark/1970s/Clark,%20H.H.%20_%20Haviland,%20S.E.%20_Comp%20and%20the%20given-new%20contract_%201977.pdf
- Clark, H. H., & Murphy, G. L. (1982). Audience Design in Meaning and Reference. *Advances in Psychology*, 9(C), 287–299. [https://doi.org/10.1016/S0166-4115\(09\)60059-5](https://doi.org/10.1016/S0166-4115(09)60059-5)
- Clark, J. K., Wegener, D. T., Sawicki, V., Petty, R. E., & Briñol, P. (2013). Evaluating the message or the messenger? Implications for self-validation in persuasion. *Personality & Social Psychology Bulletin*, 39(12), 1571–1584. <https://doi.org/10.1177/0146167213499238>
- Clark, R. A., Goldsmith, R. E., & Goldsmith, E. B. (2008). Market mavenism and consumer self-confidence. *Journal of Consumer Behaviour*, 7(3), 239–248. <https://doi.org/10.1002/cb.248>
- Clarkson, J. J., Tormala, Z. L., & Leone, C. (2011). A self-validation perspective on the mere thought effect. *Journal of Experimental Social Psychology*, 47(2), 449–454. <https://doi.org/10.1016/j.jesp.2010.12.003>
- Clarkson, J. J., Tormala, Z. L., & Rucker, D. D. (2011). Cognitive and affective matching effects in persuasion: An amplification perspective. *Personality and Social Psychology Bulletin*, 37(11), 1415–1427. <https://doi.org/10.1177/0146167211413394>
- Clarkson, J. J., Valente, M. J., Leone, C., & Tormala, Z. L. (2013). Motivated reflection on attitude-inconsistent information: an exploration of the role of fear of invalidity in self-persuasion. *Personality & Social Psychology Bulletin*, 39(12), 1559–1570. <https://doi.org/10.1177/0146167213497983>
- Clifford, S. (2019). How Emotional Frames Moralize and Polarize Political Attitudes. *Political Psychology*, 40(1), 75–91. <https://doi.org/10.1111/pops.12507>

- Clifford, S., & Jerit, J. (2013). How words do the work of politics: Moral foundations theory and the debate over stem cell research. *Journal of Politics*, 75(3), 659–671.
<https://doi.org/10.1017/S0022381613000492>
- Clifford, S., Jerit, J., Rainey, C., & Motyl, M. (2015). Moral Concerns and Policy Attitudes: Investigating the Influence of Elite Rhetoric. *Political Communication*, 32(2), 229–248.
<https://doi.org/10.1080/10584609.2014.944320>
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11), 769–771. <https://doi.org/10.1038/s41562-017-0213-3>
- Crone, D.L. (2018). An R package for annotating text data with pre-trained word embedding models. Retrieved from <https://github.com/damienchrone/embeddingtools>
- Davies, C. L., Sibley, C. G., & Liu, J. H. (2014). Confirmatory Factor Analysis of the Moral Foundations Questionnaire. *Social Psychology*, 45(6), 431–436.
<https://doi.org/10.1027/1864-9335/a000201>
- De Dreu, C. K. W., & Gross, J. (2018). Revisiting the Form and Function of Conflict: Neurobiological, Psychological and Cultural Mechanisms for Attack and Defense Within and between Groups. *Behavioral and Brain Sciences*, 42, 1–66.
<https://doi.org/10.1017/S0140525X18002170>
- DeBono, K. G., & Harnish, R. J. (1988). Source expertise, source attractiveness, and the processing of persuasive information: A functional approach. *Journal of Personality and Social Psychology*, 55(4), 541–546. <https://doi.org/10.1037/0022-3514.55.4.541>
- Delton, A. W., DeScioli, P., & Ryan, T. J. (2019). Moral Obstinacy in Political Negotiations. *Political Psychology*, 41(1), 3-20. <https://doi.org/10.1111/pops.12612>
- Department of Energy. (2018). Washington DC. Retrieved from <https://www.energy.gov/energysaver/thermostats>

- Department of Environment and Energy. (2018). Australia. Retrieved from <https://www.energy.gov.au/publications/australian-energy-update-2018>
- Dillard, J. P., & Peck, E. (2001). Persuasion and the structure of affect: Dual systems and discrete emotions as complementary models. *Human Communication Research*, 27(1), 38–68. <https://doi.org/10.1093/hcr/27.1.38>
- Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, 72(2), 144–168. <https://doi.org/10.1080/03637750500111815>
- Eagly, A. H., & Chaiken, S. (1993). The Nature of Attitudes. In C., Leone (Ed.), *The Psychology of Attitudes* (pp 1–21). Fort Worth, Tex: Harcourt Brace Jovanovich College Publishers. <https://doi.org/10.1002/mar.4220120509>
- Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, 36(4), 424–435.
- Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology*, 89(3), 257–276. <https://doi.org/10.1037/0022-3514.89.3.257>
- Echterhoff, G., Higgins, E. T., Kopietz, R., & Groll, S. (2008). How communication goals determine when audience tuning biases memory. *Journal of Experimental Psychology: General*, 137(1), 3–21. <https://doi.org/10.1037/0096-3445.137.1.3>
- Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared Reality: Experiencing Commonality With Others' Inner States About the World. *Perspectives on Psychological Science*, 4(5), 496–521. <https://doi.org/10.1111/j.1745-6924.2009.01161.x>

- Ehret, P., Van Boven, L., & Sherman, D. K. (2018). Partisan Barriers to Bipartisanship Understanding Climate Policy Polarization. *Social Psychological and Personality Science*, 9(3), 308–318. <https://doi.org/10.1177/1948550618758709>
- Ellemers, N., & van den Bos, K. (2012). Morality in Groups: On the Social-Regulatory Functions of Right and Wrong. *Social and Personality Psychology Compass*, 6(12), 878–889. <https://doi.org/10.1111/spc3.12001>
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390–405. <https://doi.org/10.1037/0022-3514.71.2.390>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G * Power 3.1. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fazio, R. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R.E Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, New Jersey: Erlbaum.
- Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, 24(1), 56–62. <https://doi.org/10.1177/0956797612449177>
- Feinberg, M., & Willer, R. (2015). From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence? *Personality and Social Psychology Bulletin*, 41(12), 1665–1681. <https://doi.org/10.1177/0146167215607842>
- Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, 32(2&3), 135–153. https://doi.org/10.1207/S15326950DP3202&3_03
- Festinger, L. (1962). Cognitive Dissonance. *Scientific American*, 207(4), 93–106.
- Festinger, L., & Carlsmith, J. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal Psychology*, 58(2), 203–210. <https://doi.org/10.1037/h0041593>

- Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science*, 19(6), 384–389.
<https://doi.org/10.1177/0963721410391246>
- Fischer, P., Jonas, E., Frey, D., & Kastenmüller, A. (2008). Selective exposure and decision framing: The impact of gain and loss framing on confirmatory information search after decisions. *Journal of Experimental Social Psychology*, 44(2), 312–320.
<https://doi.org/10.1016/j.jesp.2007.06.001>
- Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50. <https://doi.org/10.1016/j.cobeha.2015.01.010>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.
<https://doi.org/10.1016/j.tics.2006.11.005>
- Flanders, J., & Thistlethwaite, D. (1967). Effects of familiarization and group discussion upon risk taking. *Journal of Personality and Social Psychology*, 5(1), 91–97.
<https://doi.org/10.1037/h0024204>
- Fraser, C., Gouge, C., & Billig, M. (1971). Risky shifts, cautious shifts, and group polarization. *European Journal of Social Psychology*, 1(4), 493–510.
<https://doi.org/10.1002/ejsp.2420010103>
- Frimer, J. A., Brandt, M. J., Melton, Z. J., & Motyl, M. (2019). Extremists on the Left and Right Use Angry, Negative Language. *Personality & Social Psychology Bulletin*, 45(8), 1216–1231. <https://doi.org/https://doi.org/10.1177/0146167218809705>
- Frimer, J. A., Biesanz, J. C., Walker, L. J., & MacKinlay, C. W. (2013). Liberals and conservatives rely on common moral foundations when making moral judgments about influential people. *Journal of Personality and Social Psychology*, 104(6), 1040–1059.
<https://doi.org/10.1037/a0032277>

- Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3), 203–219.
- Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6), 509–525. <https://doi.org/10.1002/ejsp.2420190603>
- Garten, J., Boghrati, R., Hoover, J., Johnson, K. M., & Dehghani, M. (2016). Morality Between the Lines: Detecting Moral Sentiment In Text. *Proceedings of IJCAI 2016 Workshop on Computational Modeling of Attitudes*.
- Garten, J., Hoover, J., Johnson, K., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2018). Dictionaries and distributions: Combining expert knowledge and large-scale textual data content analysis: Distributed dictionary representation. *Behavior Research Methods*, 50(1), 344–361. <https://doi.org/10.3758/s13428-017-0875-9>
- Gascó, M., Briñol, P., Santos, D., Petty, R. E., & Horcajo, J. (2018). Where Did This Thought Come From? A Self-Validation Analysis of the Perceived Origin of Thoughts. *Personality and Social Psychology Bulletin*, 44(11), 1615–1628. <https://doi.org/10.1177/0146167218775696>
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. <https://doi.org/10.1037/a0034726>
- Gordijn, E. H., Postmes, T., & de Vries, N. K. (2001). Devil's Advocate or Advocate of Oneself: Effects of Numerical Support on Pro and Counter-attitudinal Self-Persuasion.

Personality and Social Psychology Bulletin, 27(4), 395–407.

<https://doi.org/10.1177/0146167201274002>

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013).

Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press.

<https://doi.org/10.1016/B978-0-12-407236-7.00002-4>

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and Conservatives Rely on Different

Sets of Moral Foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046. <https://doi.org/10.1037/a0015141>

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the

Moral Domain. *Journal of Personality and Social Psychology*, 101(2), 366–385.

<https://doi.org/10.1037/a0021847>

Greenberg, S., Brand, D., Pluta, A., Moore, D., & DeConti, K. (2017). Nudging Resisters

Toward Change: Self-Persuasion Interventions for Reducing Attitude Certainty.

American Journal of Health Promotion, 32(4), 997–1009.

<https://doi.org/10.1177/0890117117715295>

Greene, J. D. (2015). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science

Matters for Ethics. *Law and Ethics of Human Rights*, 9(2), 141–172.

<https://doi.org/10.1515/lehr-2015-0011>

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude

change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 147–170). New York: Academic Press.

Greenwald, A. G. (1969). The open-mindedness of the counter-attitudinal role player.

Journal of Experimental Social Psychology, 5(4), 375–388. [https://doi.org/10.1016/0022-1031\(69\)90031-6](https://doi.org/10.1016/0022-1031(69)90031-6)

- Greenwald, A. G., & Albert, R. D. (1968). Acceptance and recall of improvised arguments. *Journal of Personality and Social Psychology*, 8(1), 31–34.
<https://doi.org/10.1037/h0021237>
- Guyer, J. J., Briñol, P., Petty, R. E., & Horcajo, J. (2019). Nonverbal Behavior of Persuasive Sources: A Multiple Process Analysis. *Journal of Nonverbal Behavior*, 43, 203–231.
<https://doi.org/10.1007/s10919-018-00291-x>
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford: Oxford University Press.
- Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, 316(5827), 998–1002.
<https://doi.org/10.1126/science.1137651>
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 98–116.
<https://doi.org/10.1007/s11211-007-0034-z>
- Harmon-Jones, E., & Mills, J. (1999). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones & J. Mills (Eds.), *Science conference series. Cognitive dissonance: Progress on a pivotal theory in social psychology*. (Vol. 93, pp. 3–21). Washington: American Psychological Association.
<https://doi.org/10.1037/10318-001>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Haugtvedt, C. P., & Petty, R. E. (1992). Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of Personality & Social Psychology*, 63(2), 308–319.

- Hausmann, L. R. M., Levine, J. M., & Higgins, E. T. (2008). Communication and Group Perception: Extending the 'Saying is Believing' Effect. *Group Processes & Intergroup Relations*, 11(4), 539–554. <https://doi.org/10.1177/1368430208095405>
- Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York: Guildford Press. <https://doi.org/978-1-60918-230-4>
- Haynes, A. A., & Rhine, S. L. (1998). Attack politics in presidential nomination campaigns: An examination of the frequency and determinants of intermediated negative messages against opponents. *Political Research Quarterly*, 51(3), 691–721. <https://doi.org/10.1177/106591299805100307>
- Heesacker, M., Petty, R. E., & Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message relevant thinking. *Journal of Personality*, 51(4), 653–666. <https://doi.org/10.1111/j.1467-6494.1983.tb00872.x>
- Henningsen, D. D., & Henningsen, M. L. M. (2004). The effect of individual difference variables on information sharing in decision-making groups. *Human Communication Research*, 30(4), 540–555. <https://doi.org/10.1093/hcr/30.4.540>
- Hestres, L. E. (2015). Climate-change advocacy online: theories of change, target audiences, and online strategy. *Environmental Politics*, 24(2), 193–211. <https://doi.org/10.1080/09644016.2015.992600>
- Higgins, E. T. (1992). Achieving “shared reality” in the communication game: a social action that creates meaning. *Journal of Language and Social Psychology*, 11(3), 107–131.
- Higgins, E. T., & McCann, C. D. (1984). Social Encoding and Subsequent Attitudes, Impressions, and Memory: “Context-Driven” and Motivational Aspects of Processing. *Journal of Personality and Social Psychology*, 47(1), 26–39. <https://doi.org/10.1037/0022-3514.47.1.26>

- Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, 14(4), 363–378. [https://doi.org/10.1016/0022-1031\(78\)90032-X](https://doi.org/10.1016/0022-1031(78)90032-X)
- Hinsenkamp, D. (2018). *Extremity of a Persuasive Message Position Interacts with Argument Quality to Predict Attitude Change*. Ohio State University. Retrieved from https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu15315044530799
- 87
- Ho, S. S., & McLeod, D. M. (2008). Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication. *Communication Research*, 35(2), 190–207. <https://doi.org/10.1177/0093650207313159>
- Hoover, J., Johnson, K., Boghrati, R., Graham, J., & Dehghani, M. (2018). Moral Framing and Charitable Donation: Integrating Exploratory Social Media Analyses and Confirmatory Experimentation. *Collabra: Psychology*, 4(1), 9–27. <https://doi.org/10.1525/collabra.129>
- Horcajo, J., Paredes, B., Higuero, G., Briñol, P., & Petty, R. E. (2019). The effects of overt head movements on physical performance after positive versus negative self-talk. *Journal of Sport and Exercise Psychology*, 41(1), 36–45. <https://doi.org/10.1123/JSEP.2018-0208>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: psychological studies of opinion change*. New Haven: Yale University Press.
- Hovland, C. I., Lumsdaine, A., & Sheffield, F. D. (1949). *Experiments on mass communication*, (3rd ed.). Princeton, New Jersey: Princeton University Press.
- Hovland, C. I., & Mandell, W. (1952). An experimental comparison of conclusion-drawing by the communicator and by the audience. *Journal of Abnormal and Social Psychology*, 47(3), 581-588.

- Hovland, C.I., & Weiss, W. (1953). The influence of source credibility on communication effectiveness. *Educational Technology Research and Development*, 1(2), 635–650.
- Infante, D. A., & Rancer, A. (1993). Relations Between Argumentative Motivation, and Advocacy and Refutation on Controversial Issues. *Communication Quarterly*, 41(4), 415–426. <https://doi.org/10.1080/01463379309369902>
- Isenberg, D. J. (1986). Group Polarization. A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology*, 50(6), 1141–1151. <https://doi.org/10.1037/0022-3514.50.6.1141>
- Jacks, J., & Cameron, K. (2003). Strategies for resisting persuasion. *Basic and Applied Social Psychology*, 25(2), 145–161. https://doi.org/10.1207/s15324834basp2502_5
- Janis, I. L., & Gilmore, J. B. (1965). The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology*, 1(1), 17–27. <https://doi.org/10.1037/h0021643>
- Janis, I. L., & King, B. T. (1954). The influence of role playing on opinion change. *Journal of Abnormal Psychology*, 49(2), 211–218. <https://doi.org/10.1037/h0056957>
- Jansson, J., & Dorrepaal, E. (2015). Personal Norms for Dealing with Climate-change: Results from a Survey Using Moral Foundations Theory. *Sustainable Development*, 23(6), 381–395. <https://doi.org/10.1002/sd.1598>
- Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530(7591), 473–476. <https://doi.org/10.1038/nature16981>
- Kahneman, D. (2011). *Thinking Fast and Slow*. London, England: Penguin Books Ltd. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237.

- Kahneman, D., & Tversky, A. (1986). Prospect Theory: an Analysis of Decision Under Risk. *Econometrica*, 47(2), 263. <https://doi.org/10.2307/1914185>
- Keating, J., Van Boven, L., & Judd, C. M. (2016). Partisan underestimation of the polarizing influence of group discussion. *Journal of Experimental Social Psychology*, 65, 52–58. <https://doi.org/10.1016/j.jesp.2016.03.002>
- Kelman, H. C., & Hovland, C. I. (1953). “Reinstatement” of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology*, 48(3), 327–335.
- Kim, N. (2016). Beyond Rationality: The Role of Anger and Information in Deliberation. *Communication Research*, 43(1), 3–24. <https://doi.org/10.1177/0093650213510943>
- Kim, S. Y., Allen, M., Preiss, R. W., & Peterson, B. (2014). Meta-Analysis of Counter-attitudinal Advocacy Data: Evidence for an Additive Cues Model. *Communication Quarterly*, 62(5), 607–620. <https://doi.org/10.1080/01463373.2014.949385>
- Kim, Y. (2015). Does disagreement mitigate polarization? How selective exposure and disagreement affect political polarization. *Journalism and Mass Communication Quarterly*, 92(4), 915–937. <https://doi.org/10.1177/1077699015596328>
- King, B. T., & Janis, I. L. (1956). Comparison of the effectiveness of improvised versus non-improvised role-playing in producing opinion changes. *Human Relations*, 9(2), 177–186.
- Kline, R. B. (2015). The Mediation Myth. *Basic and Applied Social Psychology*, 37(4), 202–213. <https://doi.org/10.1080/01973533.2015.1049349>
- Knobloch-Westerwick, S., & Kleinman, S. B. (2012). Pre-election selective exposure: Confirmation bias versus informational utility. *Communication Research*, 39(2), 170–193. <https://doi.org/10.1177/0093650211400597>

- Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way. Selective exposure to attitude-consistent and counter-attitudinal political information. *Communication Research, 36*(3), 426–448. <https://doi.org/10.1177/0093650209333030>
- Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality, 46*(2), 184–194. <https://doi.org/10.1016/j.jrp.2012.01.006>
- Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The Problem with Morality: Impeding Progress and Increasing Divides. *Social Issues and Policy Review, 12*(1), 218–257. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/sipr.12045/abstract?campaign=woletoc>
- Krosnick, J. A., & Alwin, D. F. (1989). Aging and Susceptibility to Attitude Change. *Journal of Personality and Social Psychology, 57*(3), 416–425. <https://doi.org/10.1037/0022-3514.57.3.416>
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R.E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Mahwah, New Jersey: Erlbaum.
- Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.
- Kuhn, D. (2019). Why Is Reconciling Divergent Views a Challenge? *Current Directions in Psychological Science, 29*(1), 27-32. <https://doi.org/https://doi.org/10.1177/0963721419885996>
- Lamm, H., & Myers, D. G. (1978). Group-induced polarization of attitudes and behavior. *Advances in Experimental Social Psychology, 11*(C), 145–195. [https://doi.org/10.1016/S0065-2601\(08\)60007-6](https://doi.org/10.1016/S0065-2601(08)60007-6)

- Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, 34(2), 283–301. [https://doi.org/10.1016/0020-7373\(91\)90045-9](https://doi.org/10.1016/0020-7373(91)90045-9)
- Lee, J. H. (2008). Effects of news deviance and personal involvement on audience story selection: A web-tracking analysis. *Journalism and Mass Communication Quarterly*, 85(1), 41–60. <https://doi.org/10.1177/107769900808500104>
- Leidner, B., Kardos, P., & Castano, E. (2018). The Effects of Moral and Pragmatic Arguments Against Torture on Demands for Judicial Reform. *Political Psychology*, 39(1), 143–162. <https://doi.org/10.1111/pops.12386>
- Leifeld, P. (2014). Polarization of coalitions in an agent-based model of political discourse. *Computational Social Networks*, 1(7), 1–22. <https://doi.org/10.1186/s40649-014-0007-y>
- Leone, C., & Ensley, E. (1986). Self-generated attitude change: A person by situation analysis of attitude polarization and attenuation. *Journal of Research in Personality*, 20(4), 434–446. [https://doi.org/10.1016/0092-6566\(86\)90124-8](https://doi.org/10.1016/0092-6566(86)90124-8)
- Limon, M. S., Turner, M. M., & Zompetti, J. P. (2008). Informal arguing: The likelihood of providing arguments, rebuttals, refutations, and evidence in an argumentative interaction. *Argumentation and Advocacy*, 45(1), 37–48.
- Loibl, C., Cho, S. H., Diekmann, F., & Batte, M. T. (2009). Consumer self-confidence in searching for information. *Journal of Consumer Affairs*, 43(1), 26–55. <https://doi.org/10.1111/j.1745-6606.2008.01126.x>
- Love, R. E., & Greenwald, A. (1978). Cognitive Responses to Persuasion as Mediators of Opinion Change. *The Journal of Social Psychology*, 104(2), 231–241. <https://doi.org/10.1080/00224545.1978.9924065>

- Luttrell, A., Petty, R. E., Briñol, P., & Wagner, B. C. (2016). Making it moral: Merely labeling an attitude as moral increases its strength. *Journal of Experimental Social Psychology*, 65, 82–93. <https://doi.org/10.1016/j.jesp.2016.04.003>
- Luttrell, A., Phillip-Muller, A., & Petty, R. (2019). Challenging Moral Attitudes With Moral Messages. *Psychological Science*, 30(8), 1136–1150. <https://doi.org/https://doi.org/10.1177/0956797619854706>
- Manis, M., Connell, S. D., & Moore, J. C. (1974). Transmission of attitude-relevant information through a communication chain. *Journal of Personality and Social Psychology*, 30(1), 81–94. <https://doi.org/10.1037/h0036639>
- Matefy, R. E. (1972). Attitude change induced by role playing as a function of improvisation and role-taking skill. *Journal of Personality and Social Psychology*, 24(3), 343–350. <https://doi.org/10.1037/h0033725>
- Mayer, N. D., & Tormala, Z. L. (2010). “Think” Versus “Feel” Framing Effects in Persuasion. *Personality and Social Psychology Bulletin*, 36(4), 443–454. <https://doi.org/10.1177/0146167210362981>
- McCann, C. D., Higgins, E. T., & Fondacaro, R. A. (1991). Primacy and Recency in Communication and Self-Persuasion: How Successive Audiences and Multiple Encodings Influence Subsequent Evaluative Judgments. *Social Cognition*, 9(1), 47–66. <https://doi.org/10.1521/soco.1991.9.1.47>
- McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, 63(2), 326–332. <https://doi.org/10.1037/h0048344>
- McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (1st ed., pp. 191–229). New York: Academic Press.

- McGuire, W. J. (1968). Personality and attitude change: An information-processing theory. In A.G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 171–196). New York: Academic Press.
- Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, 33(2), 243–258.
<https://doi.org/10.1111/j.1467-9221.2012.00873.x>
- Miller, R. L., & Wozniak, W. (2001). Counter-attitudinal advocacy: Effort vs. self-generation of arguments. *Current Research in Social Psychology*, 6(4), 46–55.
- Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, 2(6), 389–396. <https://doi.org/10.1038/s41562-018-0353-0>
- Moorman, M., & Neijens, P. C. (2012). Political Advertising. In S. Rodgers & E. Thorson (Eds.), *Advertising theory* (pp. 297–310). New York: Taylor & Francis.
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12, 125–135. <https://doi.org/10.1037/h0027568>
- Myers, D. G. (1975). Discussion-Induced Attitude Polarization. *Human Relations*, 28(8), 699–714. <https://doi.org/10.1177/001872677502800802>
- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4), 602–627. <https://doi.org/10.1037/0033-2909.83.4.602>
- Newton, D., & Czerlinsky, T. (1974). Adjustment of attitude communications for contrasts by extreme audiences. *Journal of Personality & Social Psychology*, 30(6), 829–837.
<https://doi.org/10.1037/h0037538>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>

- Nilsson, A., & Erlandsson, A. (2015). The Moral Foundations taxonomy: Structural validity and relation to political ideology in Sweden. *Personality and Individual Differences*, 76, 28–32. <https://doi.org/10.1016/j.paid.2014.11.049>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>
- O’Rourke, H. P., & MacKinnon, D. P. (2018). Reasons for Testing Mediation in the Absence of an Intervention Effect: A Research Imperative in Prevention and Intervention Research. *Journal of Studies on Alcohol and Drugs*, 79(2), 171–181. <https://doi.org/10.15288/jsad.2018.79.171>
- Ouellette, J. A., & Wood, W. (1998). Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior. *Psychological Bulletin*, 124(1), 54–74. <https://doi.org/10.1037/0033-2909.124.1.54>
- Paredes, B., Santos, D., Briñol, P., Gomez, A., & Petty, R. E. (2019). The role of meta-cognitive certainty on the relationship between identity fusion and endorsement of extreme pro-group behavior. *Self and Identity*, 19(7), 804-824. <https://doi.org/https://doi.org/10.1080/15298868.2019.1681498>
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1162>
- Petty, R.E., & Briñol, P. (2002). Attitude change: The Elaboration Likelihood Model of persuasion. In G. Bartels & W. Nelissen (Eds.), *Marketing for sustainability: Towards transactional policy-making* (pp. 176–190). Amsterdam: IOS Press.

- Petty, R. E., & Briñol, P. (2008). Psychological Processes Underlying Persuasion: A Social Psychological Approach. *Diogenes*, 55(1), 52–67.
<https://doi.org/10.1177/0392192107087917>
- Petty, R.E., & Briñol, P. (2012). The Elaboration Likelihood Model. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 224–245). New York: SAGE Publications.
<https://doi.org/http://dx.doi.org/10.4135/9781446249222>
- Petty, R. E., & Briñol, P. (2014). The elaboration likelihood and metacognitive models of attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 172–187). New York: Guildford Press.
[https://doi.org/10.1016/s0065-2601\(08\)60214-2](https://doi.org/10.1016/s0065-2601(08)60214-2)
- Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: cognitive and meta-cognitive processes impact attitudes. *Cognition & Emotion*, 29(1), 1–26.
<https://doi.org/10.1080/02699931.2014.967183>
- Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta–Cognitive Model (MCM) of Attitudes: Implications for Attitude Measurement, Change, and Strength. *Social Cognition*, 25(5), 657–686. <https://doi.org/10.1521/soco.2007.25.5.657>
- Petty, R. E., Briñol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 318–329). New York: Guildford Press.
- Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology*, 82(5), 722–741. <https://doi.org/10.1037//0022-3514.82.5.722>
- Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of meta-cognition in social judgment. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology:*

Handbook of Basic Principles (Second, pp. 254–284). New York: Guildford Press.

Retrieved from <http://psycnet.apa.org/psycinfo/2007-11239-011>

Petty, R.E., & Brock, T. C. (1981). Thought disruption and persuasion: Assessing the validity of attitude change experiments. In R.E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 55–79). Hillsdale, New Jersey: Erlbaum.

Petty, R. E., & Cacioppo, J. T. (1979a). Effects of Forewarning of Persuasive Intent and Involvement on Cognitive Responses and Persuasion. *Personality and Social Psychology Bulletin*, 5(2), 173–176. <https://doi.org/10.1177/014616727900500209>

Petty, R. E., & Cacioppo, J. T. (1979b). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, 37(10), 1915–1926. <https://doi.org/10.1037/0022-3514.37.10.1915>

Petty, R.E., & Cacioppo, J. T. (1981). *Attitudes and Persuasion: Classic and Contemporary Approaches*. Dubuque: WC Brown Co.

Petty, R.E., & Cacioppo, J. T. (1983). Central and peripheral routes to persuasion: Application to advertising. In L. Percy & A. Woodside (Eds.), *Advertising and consumer psychology* (pp. 3–23). Lexington, Massachusetts: D.C. Health.

Petty, R. E., & Cacioppo, J. T. (1984a). Source Factors and The Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, 11(1), 668–672. <https://doi.org/10.1558/ijssl.v14i2.309>

Petty, R. E., & Cacioppo, J. T. (1984b). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, 46(1), 69–81. <https://doi.org/10.1037/0022-3514.46.1.69>

- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.
[https://doi.org/https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/https://doi.org/10.1016/S0065-2601(08)60214-2)
- Petty, R. E., & Cacioppo, J. T. (1990). Involvement and Persuasion: Tradition Versus Integration. *Psychological Bulletin*, 107(3), 367–374. <https://doi.org/10.1037/0033-2909.107.3.367>
- Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5), 847–855.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *Journal of Consumer Research*, 10(2), 135–146. <https://doi.org/10.1086/208954>
- Petty, R. E., Haugtvedt, C. P., Heesacker, M., & Cacioppo, J. T. (1995). Message elaboration as a determinant of attitude strength: Persistence and resistance of persuasion. Unpublished Manuscript.
- Petty, R.E., Ostrom, T. M., & Brock, T. C. (1981). *Cognitive Responses in Persuasion*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Petty, R. E, Priester, J. R., & Briñol, P. (2002). Mass media attitude change: Implications of the elaboration likelihood model of persuasion. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research*, (2nd ed., pp. 155–198). Hillsdale, New Jersey.
- Petty, R. E., Schumann, D. W., Richman, S.A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology*, 64(1), 5–20. <https://doi.org/10.1037/0022-3514.64.1.5>

- Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 41–72). New York: Guildford Press.
- Pew Research Center. (2008). *Americans Say They Like Diverse Communities; Election, Census Trends Suggest Otherwise*. Retrieved from <http://www.pewsocialtrends.org/2008/12/02/americans-say-they-like-diverse-communities-election-census-trends-suggest-otherwise/>
- Pew Research Center. (2014). *Political Polarization in the American Public: Growing Partisan Antipathy*. Retrieved from <https://www.people-press.org/2014/06/12/section-2-growing-partisan-antipathy/>
- Pew Research Center. (2016). *Views of candidate ‘insults,’ criticism and political divisions*. Retrieved from <http://www.people-press.org/2016/10/27/3-views-of-candidate-insults-criticism-and-political-divisions/>
- Piazza, J., & Sousa, P. (2014). Religiosity, Political Orientation, and Consequentialist Moral Thinking. *Social Psychological and Personality Science*, 5(3), 334–342. <https://doi.org/10.1177/1948550613492826>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. In *Behavior Research Methods* (Vol. 40, pp. 879–891). <https://doi.org/10.3758/BRM.40.3.879>
- Priester, J. R., & Petty, R. E. (1995). Source Attributions and Persuasion: Perceived Honesty as a Determinant of Message Scrutiny. *Personality and Social Psychology Bulletin*, 12(6), 637–654. <https://doi.org/0803973233>
- Qualtrics. (2016). Qualtrics. Provo, Utah.

- Rains, S. A. (2013). The Nature of Psychological Reactance Revisited: A Meta-Analytic Review. *Human Communication Research*, 39(1), 47–73. <https://doi.org/10.1111/j.1468-2958.2012.01443.x>
- Rancer, A. S., & Avtgis, T. A. (2006). *Argumentative and Aggressive Communication: Theory, Research, and Application*. Thousand Oaks: Sage.
<https://doi.org/10.4135/9781452225722>
- Rapp, C. (2016). Moral opinion polarization and the erosion of trust. *Social Science Research*, 58, 34–45. <https://doi.org/10.1016/j.ssresearch.2016.02.008>
- Reddit. (2016). Ask Trump Supporters.
- Reeves, B., Yeykelis, L., & Cummings, J. J. (2016). The Use of Media in Media Psychology. *Media Psychology*, 19(1), 49–71. <https://doi.org/10.1080/15213269.2015.1030083>
- Rhee, J. J., Schein, C., & Bastian, B. (2019). The What, How and Why of Moralization: A Review of Current Definitions, Methods, and Evidence in Moralization Research. *Social Psychology and Personality Compass*, 13(12), e12511.
<https://doi.org/10.1111/spc3.12511>
- Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology*, 7(4), 331–363.
<https://doi.org/10.1037/1089-2680.7.4.331>
- Rios, K., Goldberg, M. H., & Totton, R. R. (2018). An Informational Influence Perspective on (Non) conformity: Perceived Knowledgeability Increases Expression of Minority Opinions. *Communication Research*, 45(2), 241–260.
<https://doi.org/10.1177/0093650217699935>
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations. *Social and Personality Psychology Compass*, 5(6), 359–371. <https://doi.org/10.1111/j.1751-9004.2011.00355.x>

- Ryan, T. J. (2017). No Compromise: Political Consequences of Moralized Attitudes. *American Journal of Political Science*, 61(2), 409–423.
<https://doi.org/10.1111/ajps.12248>
- Sadler, O., & Tesser, A. (1973). Some Effects of Salience and Time Upon Interpersonal Hostility and Attraction During Social Isolation. *Sociometry*, 36(1), 99–112.
<https://doi.org/10.2307/2786285>
- Sagi, E., & Dehghani, M. (2014a). Measuring Moral Rhetoric in Text. *Social Science Computer Review*, 32(2), 132–144. <https://doi.org/10.1177/0894439313506837>
- Sagi, E., & Dehghani, M. (2014b). Moral Rhetoric in Twitter: A Case Study of the U.S. Federal Shutdown of 2013. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*.
- Salerno, J. M., & Peter-Hagene, L. C. (2013). The Interactive Effect of Anger and Disgust on Moral Outrage and Judgments. *Psychological Science*, 24(10), 2069–2078.
<https://doi.org/10.1177/0956797613486988>
- Schein, C., & Gray, K. (2015). The Unifying Moral Dyad: Liberals and Conservatives Share the Same Harm-Based Moral Template. *Personality and Social Psychology Bulletin*, 41(8), 1147–1163. <https://doi.org/10.1177/0146167215591501>
- Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, 22(1), 32–70.
<https://doi.org/10.1177/1088868317698288>
- Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), 103–113. <https://doi.org/10.1111/j.2041-210x.2010.00012.x>

- Sherif, M., Hovland, C. I., & Harvey, O. J. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology*, 55(2), 244–252. <https://doi.org/10.1037/h0048480>
- Shoemaker, P. J. (1996). Hardwired for news: Using biological and cultural evolution to explain the surveillance function. *Journal of Communication*, 46(3), 32–47. <https://doi.org/10.1111/j.1460-2466.1996.tb01487.x>
- Silver, J. R., & Silver, E. (2017). Why are conservatives more punitive than liberals? A moral foundations approach. *Law and Human Behavior*, 41(3), 258–272. <https://doi.org/10.1037/lhb0000232>
- Simon, H. (1979). Rational Decision-Making in Business Organizations. *The American Economic Review*, 69(4), 493–513. <https://doi.org/10.2307/1808698>
- Skitka, L. J. (2010). The Psychology of Moral Conviction. *Social and Personality Psychology Compass*, 4(4), 267–281. <https://doi.org/10.1111/j.1751-9004.2010.00254.x>
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895–917. <https://doi.org/10.1037/0022-3514.88.6.895>
- Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin*, 28(10), 1419–1429. <https://doi.org/10.1177/014616702236873>
- Smith, B. L., Lasswell, H. D., & Casey, R. D. (1946). *Propaganda, communication and public opinion*. Princeton, New Jersey: Princeton University Press.
- Smith, M. J. (1977). The effects of threats to attitudinal freedom as a function of message quality and initial receiver attitude. *Communication Monographs*, 44(3), 196–206. <https://doi.org/10.1080/03637757709390131>

- Spears, R., Lea, M., & Lee, S. (1990). De-individuation and group polarization in computer-mediated communication. *British Journal of Social Psychology*, 29(2), 121–134.
<https://doi.org/10.1111/j.2044-8309.1990.tb00893.x>
- Stoner, & Finch, J. A. (1961). *A comparison of individual and group decisions involving risk*. Massachusetts Institute of Technology. Retrieved from
http://www.researchgate.net/publication/35047969_A_comparison_of_individual_and_group_decisions_involving_risk_
- Stoner, J. A. (1968). Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology*, 4(4), 442–459.
[https://doi.org/10.1016/0022-1031\(68\)90069-3](https://doi.org/10.1016/0022-1031(68)90069-3)
- Strandberg, K., Himmelroos, S., & Grönlund, K. (2017). Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *International Political Science Review*, 40(1), 41–57.
<https://doi.org/10.1177/0192512117692136>
- Sunstein, C. (2002). The Law of Group Polarization. *The Journal of Political Philosophy*, 10(2), 175–195. <https://doi.org/10.1002/9780470690734.ch4>
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral Emotions and Moral Behavior. *Annual Review of Psychology*, 58, 345–372.
<https://doi.org/10.1146/annurev.psych.56.091103.070145>
- Tankard, M. E., & Paluck, E. L. (2017). The Effect of a Supreme Court Decision Regarding Gay Marriage on Social Norms and Personal Attitudes. *Psychological Science*, 28(9), 1334–1344. <https://doi.org/10.1177/0956797617709594>
- Tappin, B. M., & McKay, R. T. (2019). Moral polarization and out-party hostility in the US political context. *Journal of Social and Political Psychology*, 7(1), 213–245.
<https://doi.org/10.5964/jspp.v7i1.1090>

- Tausczik, Y. R., & Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Teeny, J. D., & Petty, R. E. (2018). The role of perceived attitudinal bases on spontaneous and requested advocacy. *Journal of Experimental Social Psychology*, 76, 175–185. <https://doi.org/10.1016/j.jesp.2018.02.003>
- Tesser, A. (1976). Attitude polarization as a function of thought and reality constraints. *Journal of Research in Personality*, 10(2), 183–194. [https://doi.org/10.1016/0092-6566\(76\)90071-4](https://doi.org/10.1016/0092-6566(76)90071-4)
- Tesser, A. (1978). Self-Generated Attitude Change. *Advances in Experimental Social Psychology*, 11(C), 289–338. [https://doi.org/10.1016/S0065-2601\(08\)60010-6](https://doi.org/10.1016/S0065-2601(08)60010-6)
- Tesser, A., & Conlee, M. C. (1975). Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology*, 31(2), 262–270. <https://doi.org/10.1037/h0076292>
- Tesser, A., & Cowan, C. L. (1977). Some attitudinal and cognitive consequences of thought. *Journal of Research in Personality*, 11(2), 216–226. [https://doi.org/10.1016/0092-6566\(77\)90018-6](https://doi.org/10.1016/0092-6566(77)90018-6)
- Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology*, 13(4), 340–356. [https://doi.org/10.1016/0022-1031\(77\)90004-X](https://doi.org/10.1016/0022-1031(77)90004-X)
- Tesser, A., Leone, C., & Clary, E. G. (1978). Affect control: Process Constraints versus Catharsis. *Cognitive Therapy and Research*, 2(3), 265–274. <https://doi.org/10.1007/BF01185788>
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324. [https://doi.org/10.1016/S1364-6613\(03\)00135-9](https://doi.org/10.1016/S1364-6613(03)00135-9)

- Todorov, A., Chaiken, S., & Henderson, M. (2002). The Heuristic-Systematic Model of Social Information Processing. In J.P. Dillard & M. Pfau (Eds.), *The Persuasion Handbook: Developments in Theory and Practice* (pp. 195–212). Thousand Oaks: Sage. <https://doi.org/10.4135/9781412976046.n11>
- Tormala, Z. L. (2016). The role of certainty (and uncertainty) in attitudes and persuasion. *Current Opinion in Psychology*, 10, 6-11. <https://doi.org/10.1016/j.copsyc.2015.10.017>
- Tormala, Z. L., Briñol, P., & Petty, R. E. (2004). Hidden effects of persuasion. *Advances in Consumer Research*, 31(2001), 75–76.
- Tormala, Z. L., Briñol, P., & Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of Experimental Social Psychology*, 42(5), 684–691. <https://doi.org/10.1016/j.jesp.2005.10.005>
- Tormala, Z. L., Briñol, P., & Petty, R. E. (2007). Multiple roles for source credibility under high elaboration: It's all in the timing. *Social Cognition*, 24(5), 536–552.
- Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology*, 83(6), 1298–1313. <https://doi.org/10.1037/0022-3514.83.6.1298>
- Tormala, Z. L., Petty, R. E., & Briñol, P. (2002). Ease of Retrieval Effects in Persuasion: A Self-Validation Analysis. *Personality and Social Psychology Bulletin*, 28(12), 1700–1712. <https://doi.org/10.1177/014616702237651>
- Tuller, H. M., Bryan, C. J., Heyman, G. D., & Christenfeld, N. J. S. (2015). Seeing the other side: Perspective taking and the moderation of extremity. *Journal of Experimental Social Psychology*, 59, 18–23. <https://doi.org/10.1016/j.jesp.2015.02.003>
- Tversky, A., & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>

- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458. <https://doi.org/10.1126/science.7455683>
- Tyler, T. R., & Schuller, R. A. (1991). Aging and Attitude Change. *Journal of Personality and Social Psychology*, 61(5), 689–697. <https://doi.org/10.1037/0022-3514.61.5.689>
- Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing. *Journal of Communication*, 67(5), 803–826. <https://doi.org/10.1111/jcom.12325>
- van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The Importance of Moral Construal: Moral versus Non-Moral Construal Elicits Faster, More Extreme, Universal Evaluations of the Same Actions. *PLoS ONE*, 7(11), e48693. <https://doi.org/10.1371/journal.pone.0048693>
- van Leeuwen, F., & Park, J. (2009). Perceptions of social dangers, moral foundations, and political orientation. *Personality and Individual Differences*, 47(3), 169–173. <https://doi.org/10.1016/j.paid.2009.02.017>
- van Zomeren, M. (2013). Four Core Social-Psychological Motivations to Undertake Collective Action. *Social and Personality Psychology Compass*, 7(6), 378–388. <https://doi.org/10.1111/spc3.12031>
- van Zomeren, M., Postmes, T., & Spears, R. (2012). On conviction's collective consequences: Integrating moral conviction with the social identity model of collective action. *British Journal of Social Psychology*, 51(1), 52–71. <https://doi.org/10.1111/j.2044-8309.2010.02000.x>
- van Zomeren, M., Postmes, T., Spears, R., & Bettache, K. (2011). Can moral convictions motivate the advantaged to challenge social inequality? Extending the social identity

- model of collective action. *Group Processes and Intergroup Relations*, 14(5), 735–753.
<https://doi.org/10.1177/1368430210395637>
- Vargas, P. T., Von Hippel, W., & Petty, R. E. (2004). Using Partially Structured Attitude Measures to Enhance the Attitude-Behavior Relationship. *Personality and Social Psychology Bulletin*, 30(2), 197–211. <https://doi.org/10.1177/0146167203259931>
- Vinokur, A., & Burnstein, E. (1978). Depolarization of attitudes in groups. *Journal of Personality and Social Psychology*, 36(8), 872–885. <https://doi.org/10.1037/0022-3514.36.8.872>
- Voelkel, J. G., & Feinberg, M. (2018). Morally Reframed Arguments Can Affect Support for Political Candidates. *Social Psychological and Personality Science*, 9(8), 917–924.
<https://doi.org/10.1177/1948550617729408>
- Voelkel, J. G., Brandt, M. J., & Colombo, M. (2018). I know that I know nothing: Can puncturing the illusion of explanatory depth overcome the relationship between attitudinal dissimilarity and prejudice? *Comprehensive Results in Social Psychology*, 3(1), 56–78. <https://doi.org/10.1080/23743603.2018.1464881>
- Wagner, B. C., Briñol, P., & Petty, R. E. (2012). Dimensions of Metacognitive Judgment Implications for Attitude Change. In P. Briñol & K. G. DeMarree (Eds.), *Social Metacognition* (pp. 43–61). New York: Psychology Press.
<https://doi.org/10.4324/9780203865989>
- Walter, N., & Cohen, J. (2019). When less is more and more is less: The paradoxical metacognitive effects of counterarguing. *Communication Monographs*, 86(3), 377–397.
<https://doi.org/10.1080/03637751.2019.1580378>
- Wegener, D. T., & Petty, R. E. (1997). The Flexible Correction Model: The Role of Naive Theories of Bias in Bias Correction. *Advances in Experimental Social Psychology*, 29(C), 141–208. [https://doi.org/10.1016/S0065-2601\(08\)60017-9](https://doi.org/10.1016/S0065-2601(08)60017-9)

- Wegener, D. T., Petty, R. E., & Klein, D. J. (1994). Effects of mood on high elaboration attitude change: The mediating role of likelihood judgments. *European Journal of Social Psychology*, 24(1), 25–43. <https://doi.org/10.1002/ejsp.2420240103>
- Wheeler, M. A., & Laham, S. M. (2016). What We Talk About When We Talk About Morality: Deontological, Consequentialist, and Emotive Language Use in Justifications Across Foundation-Specific Moral Violations. *Personality and Social Psychology Bulletin*, 42(9), 1206–1216. <https://doi.org/10.1177/0146167216653374>
- Wilson, T. D. (1990). Self-persuasion via self-reflection. In *Self-inference processes: The Ontario symposium*, Vol. 6. (pp. 43–67).
- Wilson, T. D., & Dunn, D. S. (1986). Effects of introspection on attitude-behavior consistency: Analyzing reasons versus focusing on feelings. *Journal of Experimental Social Psychology*, 22(3), 249–263. [https://doi.org/10.1016/0022-1031\(86\)90028-4](https://doi.org/10.1016/0022-1031(86)90028-4)
- Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology*, 47(1), 5–16. <https://doi.org/10.1037/0022-3514.47.1.5>
- Wilson, T. D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology*, 25(5), 379–400. [https://doi.org/10.1016/0022-1031\(89\)90029-2](https://doi.org/10.1016/0022-1031(89)90029-2)
- Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), 181–192. <https://doi.org/10.1037/0022-3514.60.2.181>
- Wiltermuth, S. S., & Tiedens, L. Z. (2011). Incidental anger and the desire to evaluate. *Organizational Behavior and Human Decision Processes*, 116(1), 55–65. <https://doi.org/10.1016/j.obhdp.2011.03.007>

- Wood, W., Kallgren, C. A., & Preisler, R. M. (1985). Access to attitude-relevant information in memory as a determinant of persuasion: The role of message attributes. *Journal of Experimental Social Psychology*, 21(1), 73–85.
- Wright, J. C., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance. *Personality and Social Psychology Bulletin*, 34(11), 1461–1476.
<https://doi.org/10.1177/0146167208322557>
- Yan, H., & Ford, D. (2015). *Baltimore protests turn violent; police injured*. Retrieved from <https://edition.cnn.com/2015/04/27/us/baltimore-unrest/index.html>
- Young, M. J., Tiedens, L. Z., Jung, H., & Tsai, M. H. (2011). Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition and Emotion*, 25(1), 10–21. <https://doi.org/10.1080/02699930903534105>
- Yzerbyt, V. (2016). Intergroup stereotyping. *Current Opinion in Psychology*, 11, 90–95.
<https://doi.org/10.1016/j.copsyc.2016.06.009>
- Zanna, M., & Rempel, J. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 315–334). Cambridge, United Kingdom: Cambridge University Press.
- Zimbardo, P. G. (1965). The effect of effort and improvisation on self-persuasion produced by role-playing. *Journal of Experimental Social Psychology*, 1(2), 103–120.
[https://doi.org/10.1016/0022-1031\(65\)90039-9](https://doi.org/10.1016/0022-1031(65)90039-9)
- Zuber, J. A., Crott, H. W., & Werner, J. (1992). Choice shift and group polarization: An analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology*, 62(1), 50–61. <https://doi.org/10.1037/0022-3514.62.1.50>