

THE UNIVERSITY OF CHICAGO

SELF-INTEREST AND ALTRUISM:
EXPLORING BEHAVIORS WITH MULTIPLE MOTIVATIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
MELISSA ROBIN BESWICK

CHICAGO, ILLINOIS

JUNE 2025

For my parents,
to whom I am forever grateful for investing unfathomable amounts of time, energy, and
resources into my education and personal development

“Everything we do affects other people. Every action that we take has some motivation of either being selfish or altruistic. All that adds up.”

Adam Yauch

TABLE OF CONTENTS

LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
ABSTRACT	x
1 THE USE OF EGOISTIC AND ALTRUISTIC APPEALS WHEN PERSUADING OTHERS	1
1.1 Abstract	1
1.2 Introduction	2
1.3 Relevant research and theoretical predictions	3
1.3.1 Why people may prefer sending the egoistic-only message	3
1.3.2 Why people may prefer sending the altruistic-only message	5
1.3.3 Why people may prefer sending the combined message	6
1.3.4 Egocentric anchoring and adjustment	7
1.4 The present research	8
1.5 Study 1	10
1.5.1 Method	10
1.5.2 Results	11
1.5.3 Discussion	14
1.6 Study 2a	15
1.6.1 Method	15
1.6.2 Results	17
1.6.3 Discussion	23
1.7 Study 2b	24
1.7.1 Methods	24
1.7.2 Results	25
1.7.3 Discussion	27
1.8 Study 3a	28
1.8.1 Methods	28
1.8.2 Results	30
1.8.3 Discussion	35
1.9 Study 3b	36
1.9.1 Methods	36
1.9.2 Results	37
1.9.3 Discussion	38
1.10 General Discussion	39
1.11 Conclusion	46

2	THE POWER OF CONSISTENCY: THE EFFECTS OF WORD-DEED ALIGNMENT ON BEHAVIORAL INTENTIONS	48
2.1	Abstract	48
2.2	Introduction	49
2.2.1	Credibility-enhancing displays	51
2.2.2	Other theories	53
2.2.3	Roadmap	57
2.3	Study 1	59
2.3.1	Methods	59
2.3.2	Results	61
2.3.3	Discussion	71
2.4	Study 2	72
2.4.1	Methods	72
2.4.2	Results	74
2.4.3	Discussion	80
2.5	Study 3	80
2.5.1	Methods	80
2.5.2	Results	82
2.5.3	Discussion	86
2.6	General Discussion	87
2.6.1	When inferences about beliefs do matter	90
2.6.2	Limitations and next steps	92
2.7	Conclusion	93
	REFERENCES	95

LIST OF TABLES

1.1	Proportion of participants choosing each message	11
1.2	Expectations of how convincing others will find each message vs. how personally convincing the senders find each message	12
1.3	Proportions of participants choosing each message	18
1.4	Expectations of how convincing others will find each message vs. how personally convincing the senders find each message	30
1.5	Proportions of participants' written messages categorized as egoistic, altruistic, or combining both reasons	35
1.6	Behavioral intentions compared to expected convincingness of messages and calibrated predictions	38
2.1	Solar panel second-order beliefs	62
2.2	Solar panel attitudes toward and from the ambassador	63
2.3	Disliking the ambassador predicted people's behavioral intentions more so than second-order beliefs. Among second-order beliefs, egoistic second-order beliefs had the greatest impact.	65
2.4	All three variables making up the "dislike" component (admire, annoyed, hypocrisy) had a significant negative impact on people's behavioral intentions	67
2.5	Disliking the ambassador is the greatest predictor of intentions to install solar panels for all three appeal types used	70
2.5	Disliking the ambassador is the greatest predictor of intentions to install solar panels for all three appeal types used (continued)	71
2.6	Vegetarianism second-order beliefs	75
2.7	Vegetarian ambassador attitudes	75
2.8	Disliking the ambassador predicted people's behavioral intentions more so than second-order beliefs	76
2.9	Admiration and annoyance toward the ambassador had the greatest impact on behavioral intentions among variables making up the "dislike" component	78
2.10	Disliking the ambassador predicted people's behavioral intentions more so than second-order beliefs. Among second-order beliefs, egoistic second-order beliefs had the greatest impact.	83
2.11	Among the components making up "dislike", admiration and annoyance were the largest predictors of people's behavioral intentions	84

ACKNOWLEDGMENTS

I am deeply grateful for the support I have received that has allowed me to complete this dissertation.

To members of the University of Chicago community:

To my dissertation chair, Oleg Urminsky, for patiently guiding me through the entire process of writing this dissertation. I have learned so much from you about research methods, critical thinking, communication, and how to work with data that turns out nothing like what we expected.

To my dissertation committee members, Ann McGill, Reid Hastie, and Ed O'Brien, for providing invaluable feedback at the dissertation proposal stage and advice prior to the defense. I felt supported by each of you.

To my department head, Jane Risen, for consistently offering thoughtful, sensible advice that helped me navigate the challenges that have arisen for me throughout the PhD.

To the Booth administrators, Malaina Brown, Kim Mayer, Cynthia Hillman, and Raven Davis, for your reliability and responsiveness at every stage of the program.

To the Behavioral Science faculty and students, particularly the members of my first-year cohort, for your friendliness and efforts to create a supportive environment.

To Annabelle Roberts, for your openness and acceptance towards me since we met and for accompanying me on various fun adventures.

To Scotty Behmer, for being a fantastic housemate, for the many hours of conversations in our old kitchen, and for helping me with statistics.

To other people in my life:

To Ryan Christopher Augustine Thomas, for supporting me through countless difficult situations, even ones that arose after I ignored your advice.

To my meditation teachers, Beth Upton, Daniel Ingram, Soryu Forall, North Burn, and the late S.N. Goenka, for helping me reach levels of mental clarity and peace that I did not know were possible and for exposing me to an alternative way of understanding the human mind.

To my partner Sam Sabra, for the innumerable ways you have expressed kindness, compassion, and love toward me while I have worked to complete this dissertation, from helping me manage my health to formatting this very document.

To my dad Robert Beswick, for prioritizing my education and for making other aspects of my life easier by looking out for my transportation needs, doing my taxes, and making sure I always have what I need from Costco.

To my mom Halyna Beswick, for investing so much time and energy into me ever since I was born. From reading to me and sending me to fun camps as a child to bringing me fresh green juices and home cooked meals as an adult, the ways you have supported me have given me the tools and capacity to complete this degree.

And to the countless other people and animals who have helped me grow throughout my life.

ABSTRACT

This dissertation explores topics related to behaviors that can be motivated by either self-interest or altruism. Chapter 1 investigates the strategies that laypeople use when attempting to persuade people to engage in behaviors that benefit both the self and others. When choosing between an egoistic message, an altruistic message, and messages combining both appeal types, people consistently choose to send the combined message when aiming to persuade others. We find evidence for a psychological mechanism that aligns with work on perspective-taking as egocentric anchoring and adjustment. Persuaders demonstrate egocentric anchoring, assuming that others will be convinced by the messages they themselves find convincing. They also adjust from egocentrism in systematic ways. Message senders demonstrate a “broad reach averaging” pattern, sending both egoistic and altruistic reasons in order to reach a wider range of people. Participants’ estimates about the distribution of perspectives across the population incorporate the norm of self-interest, the expectation that more people will be driven by altruistic motives than self-interested ones. We also look at the accuracy of people’s predictions about message efficacy and find that people’s expectations about which messages will be most convincing largely do not match the behavioral intentions of message recipients. We contribute to an understanding of the strategies that people without professional persuasive training use when trying to shift the behaviors of those around them.

Chapter 2 focuses on the ways in which an advocate’s word-deed alignment and misalignment impact people’s intentions to adopt the behaviors an advocate is promoting. We explore this in the case of behaviors that have not yet gained widespread acceptance. We replicate previous work in this domain finding that people are less likely to engage in the behavior being promoted after seeing an advocate for that behavior act inconsistently, either by not having solar panels installed on their own home while promoting them or by being seen eating a cheeseburger after encouraging vegetarianism. While prior work focused on the way this effect could be driven by second-order beliefs, people’s beliefs about the ambassador’s beliefs,

we find that the extent to which people dislike the ambassador is a better predictor of people's behavioral intentions. Even when including a scenario where people imagine overhearing a private conversation where the ambassador states their true beliefs, participants continue to dislike the ambassador who behaves inconsistently and report lower behavioral intentions after witnessing inconsistent behavior. The studies in this chapter offer a more nuanced perspective on the psychological mechanisms determining people's behavioral intentions in response to word-deed misalignment, highlighting that people are more likely to have a direct negative reaction to inconsistency than they are to make careful inferences about beliefs from behavior.

CHAPTER 1

THE USE OF EGOISTIC AND ALTRUISTIC APPEALS WHEN PERSUADING OTHERS

1.1 Abstract

We explore the strategies that laypeople use when attempting to persuade people to engage in behaviors that benefit both the self and others. We investigate people's choices, the psychological mechanisms underlying these choices, and the accuracy of people's predictions about which messages will be most persuasive. In three separate domains, we find that people choose to send a message combining an egoistic reason and an altruistic reason over sending a message containing only one reason. We find that people's message choices align with work on "egocentric anchoring and adjustment"; persuaders demonstrate egocentric anchoring, assuming that others will be convinced by the messages they themselves find most convincing. They then adjust away from egocentrism in the following systematic ways. Message senders demonstrate a "broad-reach-averaging" pattern, sending both egoistic and altruistic reasons in order to reach a wider range of people. When estimating the views of the population, persuaders assume that others will be more driven by self-interest than by altruism, even after controlling for what they would personally find convincing. In Study 1, we investigate people's persuasive strategies with regards to mask wearing in a pandemic, giving people the choice to send a message encouraging people to mask to protect their own health, protect others' health, or protect both their own and others' health. In Study 2a, we replicate the results of the previous study in the context of vegetarianism. In this study, we include two egoistic and two altruistic reasons, allowing us to create two-reason messages that mix egoistic and altruistic reasons or that contain the same types of reasons. We find that people's preference for two-reason messages persists, suggesting that people are not primarily choosing two-reason messages to combine egoistic and altruistic reasons. In

Study 3a, we select for participants who are genuinely motivated to convince others to avoid using single-use plastic water bottles. We replicate the findings from the previous studies and find evidence for the broad-reach-averaging hypothesis. In Study 2b and Study 3b, we look at people’s behavioral intentions after being shown the messages from Study 2a and Study 3a respectively. In both cases, the message senders’ predictions were largely inaccurate; persuaders overestimated the differences in convincingness between the messages, which were generally small. We contribute to an understanding of the strategies that people without professional persuasive training use when trying to shift the behaviors of those around them.

1.2 Introduction

Socially desirable behaviors can often be motivated in two ways: by appealing to individuals’ self-interest and by highlighting the benefits to others. Government officials have used both types of appeals while pleading with the public to engage in behaviors like social distancing and wearing masks (Centers for Disease Control and Prevention, n.d.). The opportunity to motivate others by appealing to both their egoistic and altruistic values is not unique to the COVID-19 pandemic. Installing solar panels can save people money on their own electricity bills in the long-run and protect the environment. Reducing one’s meat consumption can improve one’s own health and save animals’ lives. Limiting plastic usage can aid both one’s own health and the environment. Voting for certain policies can benefit both the voter and members of their community.

While the CDC jointly used both egoistic and altruistic appeals, it is not clear that laypeople use the same strategy. It is worth better understanding how people who have not received professional training in communication go about persuading others. Interpersonal communication between friends and family members is one of the most powerful and prevalent motivators for people to change (Berger, 2014). Moreover, people feel good when they are able to influence others (Bourgeois et al., 2009; Sommer & Bourgeois, 2010). Despite being

effective, ubiquitous, and enjoyable, this mode of communication is not well understood.

In this paper, we seek to understand the strategies that people without professional training use to persuade others when both egoistic and altruistic motives are relevant. Focusing on behaviors with the potential for self- and other-oriented gains can provide a broader, more generalizable picture of how people go about persuading others relative to focusing on behaviors that are only driven by one of these motives. Additionally, there may be less resistance to engaging in a behavior due to the self-relevant benefits relative to behaviors that are solely self-sacrificial. We explore three facets of this persuasive process. First, we explore whether people prefer sending egoistic messages, altruistic messages, or messages combining both appeal types. Second, we look at the potential psychological mechanisms underlying these choices. Third, we evaluate the message-senders' accuracy in predicting which messages are most persuasive.

1.3 Relevant research and theoretical predictions

Various theories yield different predictions when considering whether and why people would choose to send an egoistic, altruistic, or combined appeal when trying to persuade others. There is evidence for each of these theories in various contexts, but it is not clear which of these theories would best predict laypeople's persuasive strategies when both egoistic and altruistic reasons are relevant. We aim to understand how these theories apply in this new domain. Below, we lay out the various theories that we investigate in our studies. We consider the choice people might make between a message containing only an egoistic reason, a message containing only an altruistic reason, and a message combining both reasons.

1.3.1 Why people may prefer sending the egoistic-only message

Work on the norm of self-interest suggests that most people expect others to act in their own self-interest. Even when someone is personally altruistically motivated, providing direct,

experiential evidence of altruism, they tend not to expect that others will feel the same way (Miller & Ratner, 1998). This norm has been studied in a variety of contexts. When predicting how group members would apportion responsibility for joint outcomes, people assumed that others would take more credit for positive outcomes than they deserved. This “naive cynicism” characterizes the way in which people expect others, but not themselves, to be self-serving (Krueger & Clement, 1994). While this work focuses on assumptions of self-interest in the context of determining accountability, a similar pattern emerges in a business context. Managers overestimated the extent to which employees care about extrinsic incentives, like their salary and job security, relative to intrinsic incentives, like finding meaning (Heath, 1999). Given that long-time managers displayed this “extrinsic incentives bias” to the same extent as new managers, it likely represents lay theories of persuasion rather than being limited to people in managerial roles.

While the norm of self-interest has been explored when apportioning responsibility and managing employees, it is worth exploring how it applies when persuading others to engage in behaviors that could be motivated by egoistic or altruistic reasons. There are multiple avenues through which this norm can drive people to send a persuasive message centered around how a behavior will benefit the message receiver. First, this norm affects the degree to which someone expects others to be motivated by self-interest. People might genuinely expect the egoistic message to be the most effective. They may not want to dilute their message by including an appeal to altruism, if they see it as likely to be less compelling to others. Second, people may feel more comfortable using self-interested arguments because they align with social norms. Including an altruistic message could be seen as a violation of the norm of self-interest and people might expect to be judged negatively for expressing it.

1.3.2 Why people may prefer sending the altruistic-only message

Alternatively, people may choose to send only the other-oriented message. This may occur if an individual considers an altruistic action to be “moral” and treats it as sacred. People who hold strong moral beliefs feel it is taboo to mix moral and non-moral reasons (Fiske & Tetlock, 1997; Tetlock, 2003). Specifically, this theory proposes that people are averse to trading off sacred values against secular ones as engaging in this type of reasoning can feel morally corrosive (Tetlock et al., 2000). Insofar as people would categorize an altruistic reason as “moral” and as a manifestation of one of their sacred values, they will potentially feel uncomfortable mixing it with a more pragmatic, self-oriented rationale.

Regardless of whether they categorize a behavior as “moral”, many people hold the normative belief that charitable behaviors should be done for other-serving rather than self-serving reasons (Berkowitz & Daniels, 1964; Fisher et al., 2008; White & Peloza, 2009). Sending a combination of egoistic and altruistic messages may be seen as violating this norm. A potential way of avoiding negative social repercussions would be to send only one of the messages. Sending only the altruistic message is more likely to align with social norms compared to sending only the egoistic message to the extent that an action is likely to be seen as altruistic. Regardless of what reasons are explicitly given, sending only the egoistic message could still go against this norm if the domain is seen as one in which people should be prosocial.

Another possible reason people may choose to send the other-oriented message is to signal their altruism, both to others (Ariely et al., 2009) and to themselves (Bem, 1972; Bodner & Prelec, 2003). They may also intuit that others have similar concerns, and will therefore be more motivated by a message framing the behavior as an opportunity to signal altruism. Sending only the altruistic message gives the message recipient the chance to signal their own concern for others based on how they respond to the message and whether they choose to act. Including the egoistic reason may be seen as undermining this opportunity.

1.3.3 Why people may prefer sending the combined message

Rather than choosing between the egoistic-only and altruistic-only messages, people may prefer a message that combines egoistic and altruistic appeals. One possible mechanism that could drive a preference for sending the combined message is the expectation that including multiple reasons will have an additive effect, the “more-is-better” approach. This rests on the assumptions that people are evaluating the reasons in an additive way and that both reasons are generally seen positively. When people in a presentational role are viewing reasons in a piecemeal fashion and choosing which ones to send, they think additively, assuming that including all positive reasons is better. However, viewers looking at all reasons simultaneously evaluate things holistically, demonstrating an averaging pattern; including moderately good reasons can bring down the evaluation compared to including fewer reasons that are all seen very positively (Weaver et al., 2012). This “presenter’s paradox” provides one possible explanation for why message-senders may assume that including both egoistic and altruistic reasons is more persuasive, despite the effect it may have on message-receivers if one of those reasons is seen as significantly less persuasive. Similarly, prior work on impression-formation shows that people share more information than is necessary, even when incentivized to provide the amount of information they think will allow someone to form a positive impression of them (Klein & O’Brien, 2018). While we focus on behavioral intentions as opposed to impressions of the sender, this work also suggests that people may assume that sending more reasons is a better strategy. If people expect both self-oriented and other-oriented messages to be appealing to most recipients, sending both might seem like a stronger strategy than sending only one.

Message senders may also include both reasons as a means of reaching a wider audience, the “broad-reach-averaging” approach. If the sender expects different types of messages to convince different types of people, then sending both might appear to have the broadest reach, reducing the risk of sending an ineffective message. If this were the case, the sender

would have a mental model centered around averaging the perception of the reasons across all receivers. Even if the sender expects some receivers to see one of the reasons as negative, reducing the average evaluation for those individuals, including this reason could still increase the average persuasiveness overall. If participants vary in terms of which reason they find more persuasive, including two reasons increases the likelihood that they will find one appealing. This “broad-reach-averaging” hypothesis aligns with Total Unduplicated Reach and Frequency market analysis (Conjointly, n.d.). TURF analysis relies on determining how many people would like at least one option being offered, whether in reference to messages, products, or services. By maximizing the number of people who are interested in at least one option, this strategy reaches the widest possible range of customers. It is possible that lay people without formal marketing backgrounds use a similar strategy intuitively. Including a reason that some people are likely to view as negative could still be more persuasive overall when considering the impact on a group of people with diverse views.

It is worth noting that the “broad-reach-averaging” model could also explain the choice to send one of the single-reason messages. If people assume that every member of the population would only find one of the reasons convincing and not the other, then sending only the persuasive reason would be a better strategy than sending both. In this extreme case, including the second reason would not increase the number of people reached.

1.3.4 Egocentric anchoring and adjustment

Another possible psychological mechanism that could drive people’s choice of any of the three message choices is “egocentric anchoring and adjustment”. This theory claims that people start from their own point of view and then adjust when trying to take the perspective of others (Epley et al., 2004). Given that persuasion is likely to involve perspective-taking, this kind of egocentric anchoring might determine the messages people use. Particularly in the absence of other information about their target audience, people might project their own

perspective onto others. Different message-senders may make different choices among egoistic, altruistic, or combined messages, depending on which one they would personally expect to be most convinced by.

Research centered around moral persuasion in politics also indicates that people start from their own perspective when trying to persuade others. When asked to craft arguments to try to change someone’s political position, liberals and conservatives write arguments centered around their own moral convictions and fail to spontaneously consider the moral convictions of their target audience. However, when given an explicit choice between a message focused on their own values and a message that is reframed to fit the values of the target audience, people correctly identify the morally reframed message as more likely to be persuasive. Furthermore, the majority of participants choose to send the morally reframed message over the one that is aligned with their values when prompted to consider both (Feinberg & Willer, 2015, 2019). In the case of behaviors that can be motivated for egoistic or altruistic reasons, it is not yet clear whether we would see egocentrism as the dominant strategy, or if people would anticipate that others differed from themselves. We seek to build on this literature to answer the question of how people persuade others in cases that can be seen through an altruistic or self-interested lens, focusing on choices between pre-written messages.

1.4 The present research

In the following studies, we aim to understand the choices people make when trying to persuade others to engage in behaviors that can be engaged in for egoistic or altruistic reasons. In Study 1, we examine people’s strategy as message senders when trying to convince people to wear face masks early in the pandemic. We find that senders consistently prefer a combination of egoistic and altruistic messages when trying to persuade the message recipients. People’s expectations about others are driven by their own personal attitudes, aligning with work on egocentric anchoring. When considering the “adjustment” portion of “egocentric anchoring

and adjustment”, we find that people adjust their estimates of how persuasive the egoistic message will be, moving away from what they would personally find convincing and toward an expectation that others will be more likely to be motivated by self-interest.

In Study 2a, we extend these findings to the domain of vegetarianism. We build on Study 1 by including two possible egoistic reasons and two altruistic reasons, allowing us to create messages that combine the same types of messages, i.e. both egoistic or both altruistic. People’s preference for messages containing two distinct reasons persists, regardless of whether the messages combine egoistic and altruistic reasons. In Study 2b, we investigate the accuracy of people’s predictions about message efficacy by looking at recipients’ behavioral intentions in response to each message.

In Study 3a, we investigate the psychological mechanisms driving message choice when trying to motivate people to avoid using single-use plastic water bottles. We recruited a sample of “motivated persuaders,” by screening for participants who care about the environment, believe that people should avoid disposable plastic water bottles, and who say that they would be likely to try to convince others to do so. Our aim is to increase the external validity of the previous findings by using a sample of people who would genuinely be motivated to persuade others. We find that the mechanisms driving people’s choices in the previous studies also apply in a sample of motivated persuaders. In Study 3b, we evaluate the accuracy of people’s predictions. We show people who report using disposable plastic water bottles the messages and ask for their intentions to avoid disposable plastic water bottles going forward.

Across these three studies, we find evidence that the psychological model people use aligns with theories of egocentric anchoring and adjustment. People start with their own views of how convincing they would find a message and assume that others will respond similarly. They adjust away from egocentrism by assuming that others will be more likely to be motivated by self-interest. They also consider how people’s views may differ across the population, preferring combined messages with the intention of reaching a broader audience.

1.5 Study 1

1.5.1 Method

The purpose of this study was to see whether people without professional training choose to send an egoistic message, altruistic message, or messaging combining two types of reasons. Participants were presented with a scenario in which they were asked to imagine that they were a government employee responsible for promoting wearing face masks early in the COVID-19 pandemic. We did not screen for participants who were sincerely motivated to promote face masks. Instead, we asked participants to imagine that they were government employees to create a context in which they had a reason to consider various persuasive strategies, regardless of what personal views they might hold. They were then asked to make a series of choices between self-interested, other-oriented, and combined messages to indicate which message they would choose to use to convince someone to wear a face mask.

The self-interested message encouraged people to wear face masks to protect themselves: “Protect your health: Wearing a face mask when in public or around anyone you don’t live with can protect you from COVID-19. A face mask can help prevent you from breathing in virus particles. Choosing to wear a mask can lower the risk of getting sick.”

The other-oriented message was similar but was framed in terms of protecting others: “Protect others’ health: Wearing a face mask when in public or around anyone you don’t live with can protect others from COVID-19. A face mask can help prevent you from spreading virus particles. Choosing to wear a mask can lower the risk of getting others sick.”

The combined message merged the two messages: “Protect your and others’ health: Wearing a face mask when in public or around anyone you don’t live with can protect you and others from COVID-19. A face mask can help prevent you from either breathing in or spreading virus particles. Choosing to wear a mask can lower the risk of getting sick and of getting others sick.” In the combined message, the order in which the reasons were given was

counterbalanced across participants.

Participants made a series of ratings regarding how convincing they thought others would find and how convincing they personally found each of the egoistic, altruistic, and combined messages that would allow us to test for egocentrism and deviations from egocentrism. They also answered a series of questions about their own mask-wearing behavior and their general attitudes towards face masks.

Participants were recruited using Amazon Mechanical Turk in this pre-registered study (https://aspredicted.org/4G8_GX7). After removing participants who failed the attention check or had the same IP address as other responses, there were 193 valid responses out of 222 total responses.

1.5.2 Results

Main result: When given a choice between the other-oriented, self-oriented, and combined messages, 76.7% of participants choose to send the combined message (Table 1.1). In this study, we cannot comment on the accuracy of senders’ predictions because we do not have data on the actual efficacy of the messages.

Table 1.1: Proportion of participants choosing each message

Egoistic-only	Altruistic-only	Combined
11.9%	11.4%	76.7%

Process: We find evidence that the message senders expect recipients to feel similarly to how the senders feel (Table 1.2).

In a regression with clustered errors at the individual level and controlling for the message being evaluated, participants’ ratings of how personally convincing they would find each message predict how convincing they expected others to find each message, $\beta = 0.32$, $t(578) = 5.86$, $p < .001$. Additionally, expected efficacy, or how convincing they expected

Table 1.2: Expectations of how convincing others will find each message vs. how personally convincing the senders find each message

	N	Expected convincingness to others		Personal convincingness to one's self	
		Mean	SD	Mean	SD
Egoistic-only	193	5.40	1.28	5.67	1.47
Altruistic-only	193	4.75	1.45	5.72	1.56
Egoistic-altruistic combined	90	5.84	1.02	6.07	1.45
Altruistic-egoistic combined	103	5.97	1.08	5.98	1.51

others to find the message, is the main factor that predicts people's choices. Ratings of the expected convincingness for each message predict the likelihood of sending that message, *altruistic*: $\beta = 0.39$, $z(192) = 2.24$, $p = 0.03$; *egoistic*: $\beta = 0.82$, $z(192) = 3.26$, $p = 0.001$; *combined*: $\beta = 1.12$, $z(192) = 5.46$, $p < .001$.

The results are consistent with egocentrism, which predicts that when people are trying to persuade others, they begin by considering what they would find convincing themselves. Expected convincingness fully mediates the effect of personal convincingness on sending the combined message, sending the altruistic message (either on its own or in combination), and sending the egoistic message (on its own or in combination), *combined*: $a \times b = 0.05$, 95% C.I. (0.03, 0.09); *altruistic*: $a \times b = 0.01$, 95% C.I. (0, 0.02); *egoistic*: $a \times b = 0.02$, 95% C.I. (0.01, 0.04).

Expectations about convincingness of egoistic message: People expect the combined message to be the most effective, followed by the egoistic message, $t(192) = 5.04$, $p < .001$. The altruistic message on its own was expected to be the least effective and was seen as significantly less effective than the egoistic message, $t(192) = 5.26$, $p < .001$. While personal convincingness predicts expected convincingness, people are not perfectly egocentric when determining how persuasive others will find each message. Even after controlling for what participants report personally finding convincing, participants still expect the altruistic message to be less convincing than the egoistic message, *personal convincingness*: $\beta = 0.33$,

$t(383) = 5.85, p < 0.001$; *egoistic message vs. altruistic message*: $\beta = -0.67, t(383) = -5.95, p < 0.001$.

People’s expectations about the relative convincingness of each message drive their message choices. When asked to choose between only the egoistic and altruistic messages, 60% choose to send the egoistic message. Despite expectations that the altruistic message would be the least effective, about the same share of participants sent the altruistic-only message (11%) as the egoistic-only message (12%).

Preference for sending combined message: While people expect the egoistic reason to be more convincing than the altruistic reason, they tend to send the message that combines both reasons. The difference in people’s expectations about the convincingness of egoistic vs. altruistic messages predicts whether they will send the combined message. We regress whether people send the combined message vs. a single-reason message on the absolute value of the difference between the convincingness ratings of the two-reasons, the expected convincingness of the egoistic reason, and the expected convincingness of the altruistic reason. The larger the absolute value of the difference between the two reasons, the less likely people are to send both reasons, *absolute value of difference*: $\beta = -0.75, z(188) = -4.06, p < 0.001$; *expected convincingness for egoistic reason*: $\beta = -0.08, z(188) = -0.54, p = 0.59$; *expected convincingness for altruistic reason*: $\beta = -0.33, z(188) = -2.11, p = 0.03$. In sum, people are more likely to send the combined message when they see the two component messages as similarly convincing.

Altruistic attitudes and message choice: People’s personal attitudes determine which message they send. We looked at the choices and attitudes of people who strongly believe that wearing a mask is important for others’ sake, as these are the individuals who are more likely to try to persuade other people to wear a mask in their own lives. We measured support for mask-wearing for altruistic reasons using a scale comprised of the following measures: how selfish it is to not wear a mask, whether mask-wearing is a moral issue, whether people have

a moral obligation to wear masks, how much people care about saving lives, the importance of preventing people from getting sick, the importance of preventing suffering, and the degree to which they think one person wearing a mask has a positive impact on peoples' lives. These questions were coded so that positive values correspond to more positive altruistic attitudes towards wearing masks.

People with positive altruistic attitudes towards mask-wearing were more likely to send the combined message relative to those who expressed less endorsement of the other-oriented benefits of mask-wearing, $\beta = 0.32$, $z(192) = 2.28$, $p = 0.02$. Interestingly, the higher people's positive altruistic attitudes towards mask-wearing, the more likely they were to send the egoistic message, either on its own or in combination with the altruistic message, $\beta = 0.41$, $z(192) = 2.54$, $p = 0.01$. Participants with more positive altruistic attitudes may be more likely to send the self-interested message because they see it as a way of looking out for the other person. Alternatively, they may simply expect it to be more effective.

Altruistic attitudes and expectations of efficacy: Personal attitudes towards mask wearing also predict the magnitude of the difference between people's evaluations of personal convincingness and how likely they expect others to be convinced by each message. Participants with more positive altruistic attitudes toward mask wearing have a greater difference between personal and expected convincingness; more precisely, they think they are more likely than others are to care about mask-wearing, $\beta = 0.64$, $t(192) = 6.17$, $p < .001$. People who highly value mask-wearing believe that they will find messages about it more convincing than others do.

1.5.3 Discussion

When given a choice between an egoistic message, altruistic message, and a message combining egoistic and altruistic reasons, we find that people prefer to send the combined message. This contradicts the predictions of theories that suggest people would primarily choose either the

egoistic-only or altruistic-only message. Specifically, we do not find that people prefer to only appeal to others' self-interest. Additionally, participants were not so committed to the altruistic reason that they avoided pairing it with egoistic benefits. The psychological theory we find the most evidence for is egocentrism; people are largely choosing to send the combined message to others because that is the message they would personally find convincing.

However, egocentrism alone does not perfectly explain the results. Holding what participants would find personally convincing constant, they still expected the egoistic reason to be more effective than the altruistic reason. Despite expecting the egoistic reason to be more persuasive, most participants still chose to send the combined message. The smaller the difference between their evaluations of the convincingness of the egoistic and altruistic reason, the more likely people were to send the combined message. People tended to see value in both reasons, whatever their personal preferences, and to send a single message if one of the reasons was seen as significantly worse than the other. In the following study, we seek to better understand what drives peoples' preference for the combined message.

1.6 Study 2a

1.6.1 Method

In study 2a, we test whether the results of study 1 replicate in a new context: vegetarianism. Rather than using one egoistic and one altruistic reason to form messages across questions, we tested two messages of each type. This allows us to distinguish between people preferring the combined messages in study 1 either specifically because they paired an egoistic and an altruistic reasons or simply because they preferred messages with more than one reason. Looking at vegetarianism is also useful because it allows us to test whether the findings in study 1 generalize to a behavior that is practiced less commonly than mask-wearing. Mask-wearing was very common at the time of Study 1: 81.3% of participants reported that

they always wore a mask when around people from outside their household, 16.1% reported that they sometimes wore a mask when around people from outside their household, and only 2.6% reported that they rarely or never wore a mask around people from outside their household. In contrast, only 13.2% of participants were vegetarian or vegan in Study 2a.

General population participants (i.e. mostly non-vegetarians) were shown a scenario in which they imagined that they worked in marketing for a company selling vegetarian foods. Asking participants to imagine that they worked in marketing served to create a context in which they would need to consider what might convince people to go vegetarian, regardless of their own personal dietary choices. They were then asked to choose between a series of messages with the goal being to convince people to try vegetarianism. The messages were composed of self-interested reasons for vegetarianism (improving health or saving money) and other-oriented reasons for vegetarianism (helping animals or saving the environment). The messages read as follows:

Health: “Going vegetarian can help you improve your health. A vegetarian diet can improve heart health and reduce the risk of obesity and various types of cancer. Choosing to stop eating meat can make you healthier and extend your life.”

Money: “Going vegetarian can help you save money. At both grocery stores and restaurants, vegetarian foods are often cheaper than meat. Choosing to stop eating meat can cost less than eating meat.”

Animal Welfare: “Going vegetarian can help you improve animals’ lives. In factory farms, animals are subject to overcrowding, exposure to disease, and high levels of stress up until they are slaughtered. Choosing to stop eating meat is a way to stop supporting this industry.”

Environment: “Going vegetarian can help you save the environment. A vegetarian diet can reduce carbon and greenhouse gas emissions, helping to slow down

climate change. Choosing to stop eating meat is a way to help the planet.”

Participants saw a series of questions containing three choice options: two of the single-reason messages and one message that combined those two reasons. They saw these for all six possible two-reason combinations of the four messages (Health, Money, Animal Welfare, and Environment).

They then made a series of choices between three combined, double-reason messages. In each of these questions, they saw the message containing two egoistic reasons (health-money), the message containing two altruistic reasons (animals-environment), and one of the messages containing one altruistic and one egoistic reason (health-animals, health-environment, money-animals, or money-environment).

Participants also rated how compelling they found the messages. For the component single-reason messages (health, money, animals, environment), participants rated how convincing they thought others would find each message and how convincing they personally found each message. They also answered questions about their personal attitudes towards vegetarianism.

Participants were recruited using Amazon Mechanical Turk. This study was pre-registered (https://aspredicted.org/X9H_W7Q). After removing participants who failed the attention check or had the same IP address as other respondents, there were 205 valid responses out of 221 total collected responses.

1.6.2 Results

Main result: Consistent with study 1, people were most likely to choose the combined messages. Across all six questions that included choices between two messages containing single reasons for vegetarianism and a message containing a combination of those reasons, participants chose the combined message 57% of the time (Table 1.3), significantly more than half, $\chi^2(1, N = 820) = 27.23, p < 0.001$. People preferred the messages that contained two reasons to messages that contained only a single reason, regardless of whether the combined

Table 1.3: Proportions of participants choosing each message

	Animals	Health	Combined
Animals-Health	7.3%	41.0%	51.7%
	Money	Health	Combined
Money-Health	6.8%	25.4%	67.8%
	Environment	Health	Combined
Environment-Health	4.9%	37.1%	58.0%
	Animals	Money	Combined
Animals-Money	16.1%	35.1%	48.8%
	Environment	Money	Combined
Environment-Money	14.1%	30.7%	55.1%
	Animals	Environment	Combined
Animals-Environment	16.6%	20.0%	63.4%

message included only egoistic, only altruistic, or a combination of egoistic and altruistic reasons.

When looking only at the questions between three options, an egoistic reason, an altruistic reason, and the combination of the two reasons, 53% chose the combined message. This was significantly more than $\frac{1}{3}$, which is what would be expected by chance, $\chi^2(1, N = 820) = 147.9, p < 0.001$. People preferred the messages with two reasons, even when they combined both an egoistic and an altruistic reason.

Process: Participants' views of how persuaded they would be by each message predicted their expectations of how likely others were to be persuaded. The more personally convincing participants found the message, the more persuasive they thought others would find the message, $\beta = 0.36, t(819) = 9.06, p < .001$. People's evaluations of message persuasiveness to others differed as a function of their own personal beliefs regarding each of the moral messages, *animals*: $\beta = 0.41, t(203) = 4.75, p < .001$; *environment*: $\beta = 0.29, t(203) = 3.78,$

$p < .001$. We also measured the degree to which participants thought vegetarianism would improve health and save people money. Participants’ beliefs about the egoistic reasons did not predict how convincing they expected these reasons to be for others, *health*: $\beta = 0.04$, $t(203) = 0.71$, $p = 0.48$; *money*: $\beta = 0.09$, $t(203) = 1.06$, $p = 0.29$. People who believed that the egoistic messages accurately represented the benefits of vegetarianism were not more likely to expect them to be persuasive to others. Personal endorsement predicted people’s evaluations of persuasiveness more for the altruistic messages than for the egoistic ones.

In turn, expected convincingness predicted message choice. When looking at whether people sent an other-oriented message, either on its own or in a combined message, evaluations of that message’s efficacy predicted message choice, $\beta = 0.39$, $z(819) = 6.37$, $p < .001$. The same is true for the self-oriented message, $\beta = 0.38$, $z(819) = 5.86$, $p < .001$.

As in study 1, expected convincingness fully mediates the effect of how personally convincing someone finds the egoistic message on whether they choose to send that message, either on its own or as part of the combined message, $a \times b = 0.03$, 95% C.I. (0.02, 0.03). However, expected convincingness only partially mediates the effect of personal convincingness on whether someone sends the moral message, $a \times b = 0.04$, 95% C.I. (0.03, 0.04). This differs somewhat from the results in study 1, where this effect was fully mediated. The partial mediation suggests that there may be an omitted mediator that would further explain the relationship between personal convincingness and choice of the moral message in this context.

Expectations about egoistic message: Participants generally expected others to be self-interested. As in study 1, people expected others to be more likely to be persuaded by the self-oriented rather than the other-oriented messages, $\beta = 0.74$, $t(819) = 4.35$, $p < .001$. While the average rating of convincingness to others of the “health” message is significantly higher than the “money” message, the “money” message was expected to be more effective for others than either the “animals” or “environment” messages, *money* > *animals*: $t(204) = 4.35$, $p < .001$; *money* > *environment*: $t(204) = 4.34$, $p < .001$.

While people’s evaluations of how convincing they would personally find each message predict their expectations of how convincing others would find each message, there are some differences. Replicating the results from study 1, participants expected the egoistic message to be more convincing than the altruistic message, regardless of which message they report as more personally persuasive. Expected convincingness to others is jointly predicted by both personal convincingness of the message and whether the message is egoistic or altruistic: *personal convincingness*: $\beta = 0.37$, $t(817) = 9.35$, $p < 0.001$; *egoistic message vs. altruistic message*: $\beta = -0.97$, $t(817) = -8.05$, $p < 0.001$.

When considering the proportion of people who send each type of message (egoistic or altruistic), either on its own or as part of a combined message, more people choose to send an egoistic message (76%) relative to an altruistic message (59%). People are unlikely to send the animal-only (12%) or environment-only messages (10%), regardless of their personal attitudes. In fact, their personal views do not predict their likelihood of sending these altruistic messages, *animals*: $\beta = 0.11$, $z(409) = 0.87$, $p = 0.39$; *environment*: $\beta = 0.17$, $z(409) = 1.40$, $p = 0.16$. Even people with strong altruistic attitudes chose to include the self-interested message.

We also asked a series of questions in which participants chose between the following three message options: animals and environment (altruistic), health and money (egoistic), and one of the four possible combinations that each include one egoistic and one altruistic message. Most people (63%) choose to send the message containing two egoistic reasons. When we include only double-reason messages, people expect others to be most motivated by self-interest. This effect is moderated by people’s altruistic attitudes. Even though the message with two egoistic reasons was consistently the most popular, it was less popular among those with more positive altruistic attitudes for both animals and for the environment, *animals*: $\beta = -0.37$, $z(409) = -3.46$, $p = 0.001$; *environment*: $\beta = -0.24$, $z(409) = -2.39$, $p = 0.002$.

Preference for two-reason messages: As in Study 1, participants’ preference for sending

the combined message is predicted by the difference in perceived convincingness between the two reasons included in the message. We conducted a regression analysis looking only at the questions where participants were choosing between an egoistic-only message, an altruistic-only message, and a combined egoistic-altruistic message, clustering standard errors at the level of the individual. When controlling for the expected convincingness of the egoistic and altruistic messages, the absolute value of the difference in convincingness between the messages is negatively related to peoples' decision to send the two-reason message, *absolute value of difference*: $\beta = -0.25$, $z(817) = -3.37$, $p < 0.001$; *expected convincingness for egoistic reason*: $\beta = 0.15$, $z(817) = 2.47$, $p = 0.01$; *expected convincingness for altruistic reason*: $\beta = 0.20$, $z(817) = 3.00$, $p = 0.002$. When looking at the questions where participants chose between messages with the same type of reason (either both egoistic or both altruistic), the result is the same: the larger the difference in convincingness, the less likely people are to send the two-reason message, $\beta = -0.49$, $z(408) = -5.07$, $p < 0.001$.

Preference for two-reason messages with congruent reasons: Overall, people preferred to send messages with two reasons rather than just one. However, the preference for two reasons is stronger when the messages are either both altruistic or both egoistic, relative to when egoistic and altruistic reasons are combined. When controlling for both the personal and general convincingness of each message, however, we find that people choose the combined message slightly more frequently when it includes two of the same types of reasons as opposed to two different types of reasons, $\beta = -0.33$, $z(1229) = -3.09$, $p < 0.01$. Therefore, the preference for congruent reasons seems to be attributable to heterogeneity in what senders find convincing, and their choices reflect an additional preference for sending either both moral or both egoistic reasons, all else equal.

Altruistic attitudes and message choice: Participants' scores on the animal attitudes scale consisted of the average of the following measures: how selfish it is to kill animals for food, whether eating meat is a moral issue, the importance of saving animals' lives, the

importance of preventing animals from suffering, and whether one person going vegetarian has a positive impact on animals' lives. The environmental attitudes scale included similar questions: whether a vegetarian diet can help the environment, if producing meat negatively impacts the environment, whether going vegetarian to prevent climate change is a moral issue, the importance of avoiding meat to prevent environmental harm from methane, the importance of protecting the environment through what you eat, and whether one person going vegetarian positively impacts the environment. For both scales, the questions were coded so that higher scores corresponded to more positive attitudes.

Participants' existing altruistic attitudes predicted their choice of message. People with positive attitudes toward animals were more likely to pick the combined message relative to those with low animal attitudes, $\beta = 0.20$, $z(1639) = 2.10$, $p < 0.04$. The same pattern holds regarding positive environmental attitudes, $\beta = 0.23$, $z(1639) = 2.62$, $p < 0.01$. This is consistent with what we found in study 1. This may occur because participants with more positive altruistic attitudes tend to view both the other-oriented and the self-oriented reasons more favorably than those with lower altruistic attitudes. We do not find support for theories that predict that those with strong moral attitudes would be averse to combining an egoistic reason with a moral reason.

People are more likely to send the altruistic message, either on its own or in combination with the egoistic message, when they have more positive altruistic attitudes, *animals*: $\beta = 0.26$, $z(409) = 2.42$, $p = 0.02$; *environment*: $\beta = 0.32$, $z(409) = 3.27$, $p = 0.001$. In contrast, participants' altruistic attitudes did not affect their likelihood of sending the altruistic message in study 1. This discrepancy is likely explained by the fact that almost everyone in study 1 sent the altruistic reason, either alone or with the egoistic reason, whereas there was more variation in people's message choices in study 2.

Altruistic attitudes and expectations of efficacy: As in study 1, those with more positive altruistic attitudes have a larger difference between their evaluations of personal convincingness

and their expectations regarding how convincing others will find each message, *animals*: $\beta = 0.57$, $t(203) = 5.82$, $p < .001$; *environment*: $\beta = 0.56$, $t(203) = 6.52$, $p < .001$. For both the animal message and environment message, those with higher altruistic attitudes evaluate their own likelihood of being convinced by the message as significantly higher than others' likelihood. While those with more positive altruistic attitudes are more likely to send the altruistic message, they also seem to recognize that they might care more about animals and the environment relative to other individuals in the population.

1.6.3 Discussion

Participants preferred to send the messages containing two reasons over the messages containing a single reason. Replicating the results from study 1 using vegetarianism rather than mask-wearing, preferences for combined messages persisted even in a context where the behavior is not commonly practiced. Furthermore, the messages with two reasons were more appealing than the single-reason messages, regardless of whether they included only egoistic, only altruistic, or a combination of egoistic and altruistic reasons. While the preference for the two-reason message held in general, this preference was stronger when the messages were either both egoistic or both altruistic.

We find more evidence for egocentrism; people are more likely to include the reasons they personally find persuasive. As in Study 1, participants are not entirely egocentric; they expect the egoistic messages to be more convincing than the altruistic messages, even when controlling for what they would personally find convincing. Despite expectations that the egoistic message will be more convincing, people still choose to send the two-reason messages; when the difference between the two messages was small, people were more likely to include both. While people have strong intuitions about which messages are most likely to be more effective, it is not clear whether these predictions are accurate. In the following study, we compare people's predictions to actual behavioral intentions in response to the

various messages.

1.7 Study 2b

The purpose of this study is to evaluate the accuracy of people’s expectations about message efficacy. We look at people’s intentions to try vegetarianism after receiving each of the single-reason and double-reason messages from study 2a. We chose to test the influence of messages on intentions for vegetarianism rather than mask-wearing, because the vast majority of participants (97.4%) in study 1 indicated that they either always or sometimes wore a mask when around people from outside their household. We anticipated that, at the time of the study, people would be strongly in favor of mask-wearing, regardless of which message they were shown. In contrast, vegetarianism is much less common (13.2% of participants in Study 2), so we expected to see more variation in behavioral intentions.

1.7.1 *Methods*

Participants were recruited using Amazon Mechanical Turk. We pre-registered this study (https://aspredicted.org/MJK_TGX). Of the 792 collected responses, 759 were valid after removing participants who failed the attention check or had repeated records with the same IP address. Using a between subjects design, each participant saw one of 10 possible messages presenting one or two reasons to try vegetarianism. The messages included each of the four single-reason messages (health, animals, money, environment) in addition to the six possible ways of combining these four messages into two-reason messages. The order of the reasons within the combined messages was counterbalanced across participants. After seeing each message, participants used a 7-point Likert scale to answer the question, “How likely would you be to go vegetarian?”.

1.7.2 Results

Combined vs. single messages: Across study 1 and study 2a, people consistently chose to send messages containing two reasons as opposed to sending messages that only contained one reason. In most cases in study 2b, people’s behavioral intentions after being shown a message do not significantly differ between the combined messages and the single messages. The only instance in which there is a significant difference is when comparing health-only to health-money, and this result is in the opposite direction of what most participants in Study 2a predicted. Study 2b participants reported a higher likelihood of going vegetarian in response to the health message than the health-money message, $t(68.5) = 2.04, p = 0.045$.

Egoistic messages: When looking at Likert scale ratings of the single-reason messages, participants in study 2a expected the egoistic messages to be more effective than the altruistic messages. People were accurate in predicting that the health message will be the most effective. However, when looking at behavioral intentions in Study 2b, health is not significantly different from the animal-only and environment-only messages, *animals*: $t(149.3) = 1.50, p = 0.14$; *environment*: $t(148.0) = 1.16, p = 0.25$. Participants in study 2a overestimated the extent to which health would be more effective than the altruistic messages. Furthermore, the money-only message was the least effective. Though not significantly different from the animal or environmental message, the money message was significantly less effective than health, $t(146.4) = 2.53, p = 0.01$.

In study 2a, participants’ choices in the questions containing only double-reason messages also indicated that they expected others to be more motivated in response to the egoistic messages. When given the choice of health-money (egoistic-only), animals-environment (altruistic-only), and a message mixing egoistic and altruistic reasons, participants chose the egoistic-only message 63% of the time, significantly more than half, $\chi^2(1, N = 820) = 58.5, p < 0.001$. They chose the altruistic-only message 20% of the time and chose a mixed egoistic-altruistic message only 16% of the time. Among participants who chose the altruistic-only

message or the egoistic-only message, more than half (76%) chose health-money, $\chi^2(1, N = 685) = 182.9, p < 0.001$. However, these beliefs about relative effectiveness were not correct. When looking at behavioral intentions in 2b, the opposite was found: animals-environment led to higher intentions of going vegetarian relative to health-money, $t(148.9) = 2.10, p = 0.037$. While participants in Study 2a expected others to be more affected by the egoistic reasons (both for single-reason messages and double-reason messages), this did not consistently occur in Study 2b.

Comparing intentions in 2a to 2b: We also compare people’s evaluations of how likely they would personally be to go vegetarian in response to each message across studies 2a and 2b. In 2a, participants were asked “How likely are these messages to convince you to go vegetarian?” and then were presented with the four single-reason messages. In 2b, they were asked “How likely would you be to go vegetarian?” There are several similarities between the results of Study 2a and Study 2b, suggesting that people’s estimates were accurate in some instances. Predicted personal convincingness in Study 2a is similar to measured effectiveness in Study 2b. In both studies, the health-only message leads to higher behavioral intentions than money-only, 2a: $t(204) = 5.22, p < .001$; 2b: $t(146.4) = 2.53, p = 0.012$. Even among egoistic reasons, there is significant variation in efficacy. Money is not significantly different from animals or the environment, 2a: *animals*: $t(204) = -1.00, p = 0.32$; *environment*: $t(204) = 0.81, p = 0.42$; 2b: *animals*: $t(149) = -1.05, p = 0.30$; *environment*: $t(148.0) = -1.43, p = 0.16$. This demonstrates that egoistic appeals are not universally more effective than altruistic appeals, regardless of within-subjects or between-subjects evaluation mode.

There were also results across these two studies that highlight the ways in which people’s beliefs about message efficacy are mistaken. In study 2a, people reported that they personally found the animal-only message to be more convincing than the environment-only message, $t(204) = 2.96, p = 0.003$. They also evaluated the personal convincingness of the health message more highly than the animal and environment messages, *animals*: $t(405.4) = 3.03,$

$p = 0.003$; *environment*: $t(405.4) = 4.72$, $p < .001$. While the directions of these findings were the same as in 2a, in 2b the differences were much weaker, and there were no significant differences between the personal intention to try vegetarianism when shown the health-only, animals-only, or environment-only messages. Because participants in 2a saw all the messages, whereas participants in 2b each only saw one message, the differences in these messages may have been more evaluable in Study 2a.

Note that while participants evaluated all messages side-by-side in 2a (within subjects), each participant saw only one message in 2b (between subjects). It is possible that the inconsistencies between message choices in Study 2a and response to messages in Study 2b could be explained by differences in evaluability (Shen et al., 2012). However, it is important to note that this difference in evaluation mode is realistic: senders typically consider multiple messages whereas receivers focus on the one message they did receive. Additionally, in 2b, participants were not answering questions about what they thought others would find convincing.

1.7.3 Discussion

In Study 2a, people consistently choose to send combined messages rather than messages that contain only one reason. In this study, we find small differences in people’s behavioral intentions in response to the different types of messages. In most instances, people’s intentions do not significantly vary between the single-reason and combined two-reason messages. When choosing between which messages to send, people overpredict the degree to which effectiveness will differ. It is worth noting that, overall, there is very little variation in people’s behavioral intentions in study 2b. When evaluating how likely they were to go vegetarian in response to the single-reason and double-reason messages, people’s responses range from 2.80/7 (money-only) to 3.56/7 (health-only). The small differences limit our ability to draw strong conclusions. While participants in Study 2a had clear intuitions about which messages are most effective,

they were likely overestimating the differences between the efficacy of the messages, insofar as behavioral intentions serve as a proxy for behavior.

1.8 Study 3a

1.8.1 *Methods*

The primary purpose of this study was to investigate whether the previous results would replicate in the case of a sample of people who were actually motivated to persuade others to change their behavior. Participants were recruited using Prolific in this pre-registered study (<https://aspredicted.org/h4zj-jbpd.pdf>). In this study, we aimed to understand the choices and psychological mechanisms for people who would be sincerely motivated to convince others. We used a built-in prescreener from Prolific and only included participants who responded with 4 or 5, with 5 being “Very Concerned”, to the following question: “Generally speaking, how concerned are you about environmental issues?” Within the analyses, we only included data from participants who responded “Yes” to “Do you think people should avoid using single-use plastic water bottles?” and with 5, 6, or 7 (7 = Extremely likely) to “How likely would you be to try to convince someone to avoid single-use plastic water bottles?” After removing participants who failed the attention check or had the same IP address as other responses, we were left with a sample of 520 people who were both aligned with the issue at hand and were already inclined to try to persuade others.

Participants were asked to make choices between self-interested, other-oriented, and combined messages as in the previous studies. These were the messages that participants were shown:

Egoistic: “Protect your health: Avoiding single-use plastic water bottles can protect you from directly drinking small plastic particles. These particles can build up in your body, potentially causing cancer and heart problems. Avoiding

plastic water bottles can lower your risk of getting sick.”

Altruistic: “Protect others’ health: Avoiding single-use plastic water bottles can prevent bottles from polluting bodies of water, where small plastic particles can spread into drinking water. These particles can build up in people’s bodies, potentially causing cancer and heart problems. Avoiding plastic water bottles can lower your risk of getting others sick.”

Combined (with order of reasons counterbalanced): “Protect your and others’ health: Avoiding single-use plastic water bottles can protect you from directly drinking small plastic particles. It can also prevent bottles from polluting bodies of water, where small plastic particles can spread into drinking water. These particles can build up in your and other people’s bodies, potentially causing cancer and heart problems. Avoiding plastic water bottles can lower your risk of getting sick and getting others sick.”

In addition to the questions that were asked in previous studies, we added new questions in order to better understand the psychological mechanisms driving people’s choices. We asked participants what proportion of message recipients would likely be convinced by at least one of the single-reason messages. We also asked them to estimate what proportion of message recipients would be convinced by the combined message. This allowed us to evaluate whether participants preferred the combined message because they thought that including both reasons made the message stronger (the “more-is-better” hypothesis) or because they were trying to reach as many people as possible who would find at least one of the reasons compelling (the “broad-reach-averaging” hypothesis).

We also asked half the participants to write their own persuasive message before seeing any of the pre-written messages. This gave us the opportunity to evaluate whether people who are genuinely motivated to persuade others spontaneously prefer to combine egoistic and altruistic reasons when generating their own messages (i.e., as opposed to only when

provided with suggestions).

1.8.2 Results

Main result: As in the previous studies, a majority of participants (81.6%) chose the combined message over the other-oriented (5.7%) or self-interested messages (12.7%).

Process: As in prior studies, personal convincingness predicted expected convincingness (Table 1.4). In a regression controlling for the message being evaluated with clustered standard errors at the individual level, participants' evaluations of how personally convincing they would find each message predict how convincing they expected others to find each message, $\beta = 0.56$, $t(1517) = 15.06$, $p < .001$.

Table 1.4: Expectations of how convincing others will find each message vs. how personally convincing the senders find each message

	Expected convincingness to others		Personal convincingness to one's self	
	Mean	SD	Mean	SD
Egoistic-only	5.72	1.05	6.05	0.95
Altruistic-only	4.74	1.48	5.19	1.27
Egoistic-altruistic combined	6.18	0.99	6.29	0.92

For the egoistic-only and combined messages, ratings of expected convincingness predicted the likelihood of sending those messages, *egoistic*: $\beta = 0.41$, $z(505) = 2.76$, $p = 0.006$; *combined*: $\beta = 0.91$, $z(505) = 7.64$, $p < 0.001$. There was not a significant effect when looking at the altruistic-only message, possibly because fewer than 6% of participants chose to send this message. While not significant, the effect did trend in the same direction as the previous results, *altruistic*: $\beta = 0.21$, $z(505) = 1.47$, $p = 0.14$. These results are consistent with egocentric anchoring as a determinant of message choice.

We performed mediation analyses evaluating whether expected convincingness mediates the effect of personal convincingness on message choice. We found evidence for partial

mediation when looking at people’s choice to send the combined message or to include the altruistic reason, either in its own message or as part of the combined message, *combined*: $a \times b = 0.06$, 95% C.I. (0.03, 0.09); *altruistic*: $a \times b = 0.01$, 95% C.I. (−0.01, 0.02). When looking at people’s choice to send the egoistic message, either on its own or in combination with the altruistic message, we did not find evidence for this mediation, *egoistic*: $a \times b = 0$, 95% C.I. (−0.01, 0.01). Given that 94% of participants included the egoistic message in their choice of message, there was likely not enough variation for this mediation to be meaningfully informative.

Expectations about convincingness of egoistic message: People expected the combined message to be most effective, followed by the egoistic-only message, $t(506) = 8.67$, $p < 0.001$. They expected the altruistic message to be the least effective. The altruistic message was seen as significantly less likely to persuade others than the egoistic message, $t(506) = -14.99$, $p < 0.001$. When asked to choose between sending an egoistic-only or altruistic-only message, 79.6% of participants chose the egoistic message.

Participants still expected the altruistic message to be less convincing than the egoistic message after controlling for personal convincingness, *personal convincingness*: $\beta = 0.55$, $t(519) = 13.79$, $p < 0.001$; *egoistic message vs. altruistic message*: $\beta = -0.50$, $t(519) = -6.94$, $p < 0.001$. In the “adjustment phase” of egocentric anchoring and adjustment, one way in which people are adjusting their predictions is based on the assumption that people tend to be more driven by self-benefitting reasons rather than by altruistic ones.

Preference for sending combined message: While people expect the egoistic reason to be more convincing than the altruistic reason, they still strongly prefer to send the message combining both types of reasons. We regressed people’s choice of the combined message on the absolute value of the difference between the convincingness ratings of the egoistic and altruistic messages, controlling for the convincingness ratings. As in the previous studies, people were less likely to send the combined message when the absolute value of the difference

was larger. While the effect was weaker than in previous studies, it was directionally the same, *absolute value of difference*: $\beta = -0.29$, $z(503) = -1.89$, $p = 0.06$; *expected convincingness for egoistic reason*: $\beta = -0.15$, $z(503) = -1.03$, $p = 0.31$; *expected convincingness for altruistic reason*: $\beta = -0.08$, $z(503) = -0.56$, $p = 0.57$.

Altruistic attitudes and message choice: We investigated the choices and attitudes of people who strongly believe that avoiding single-use plastic water bottles is important for altruistic reasons. While everyone in the sample was genuinely motivated to persuade others on this topic, there was still variation in terms of the strength of people’s beliefs. We adapted the mask-wearing scale from Study 1 to measure people’s attitudes toward plastic bottle usage, including the following measures: how selfish it is to use single-use plastic water bottles, whether plastic water bottle usage is a moral issue, whether people have a moral obligation to avoid using plastic water bottles, how much people care about saving lives, the importance of preventing people from getting sick, the importance of preventing suffering, and the degree to which they think one person avoiding single-use plastic water bottles has a positive impact on people’s lives. Higher values correspond to more positive altruistic attitudes.

As in the previous studies, people with positive altruistic attitudes toward avoiding single-use plastic water bottles were more likely to choose the combined message over the single-reason messages, $\beta = 0.54$, $z(500) = 3.58$, $p < 0.001$. In the mask-wearing study, we had found that positive altruistic attitudes toward mask-wearing predicted people’s likelihood of sending the egoistic message, either on its own or in combination with the altruistic message. We also found this in the current study, $\beta = 0.51$, $z(500) = 2.12$, $p = 0.03$. These results provide further evidence that people who hold strong moral attitudes are still in favor of appealing to others’ self-interest, despite that not being their primary motivation.

Psychological mechanisms driving message choice: In addition to using the measures that we had included in the prior studies, we added new measures to better understand the psychological mechanisms driving message choice, particularly in the case of people who are

genuinely motivated to persuade others. Specifically, we wanted to measure beliefs about heterogeneity in message effectiveness across recipients to test the broad-reach-averaging model.

One set of measures was intended to understand how participants considered the distribution of message convincingness across the population of possible recipients. We asked participants to estimate what proportion of people would be convinced by either the egoistic or altruistic message, by only the egoistic message, only the altruistic message, or neither message. Participants expected 36.2% of people to be convinced by seeing either the egoistic-only or altruistic-only messages, 35.3% to be convinced only by the egoistic message, 18.7% to be convinced only by the altruistic message, and 9.8% to be convinced by neither message. This means they expected 90.2% of people to be convince-able (i.e., by at least one of the single-reason messages), but only 71.5% (i.e., $35.3\% + 36.2\%$) to be convinced by the strongest single message (the egoistic message). In the second question, we also asked participants to estimate what proportion of people would be convinced by the combined message. Participants estimated that 70.8% of people would be convinced by the combined message.

On average, participants expected a greater proportion of participants to be convinced by at least one of the single-reason messages (90.2%) than by the combined message (70.8%), $t(482) = 20.96$, $p < 0.001$, 95% CI [17.55, 21.18]. To understand the number of participants who demonstrated this pattern, we put people into two categories: those who expected more people to be convinced by at least one of the single-reason messages than by the combined message (the “broad-reach-averaging” group), and those who expected more people to be convinced by the combined message than by at least one of the single-reason messages (the “more-is-better” group). We found that 83.0% of participants fit into the “broad-reach-averaging” group, expecting more people to be convinced by at least one of the single-reason messages than by the combined message. This suggests that most participants were aiming

to maximize the number of people who received a message with at least one reason that they would find persuasive. If people were expecting the two reasons to work better together than only using one, we would have expected people’s estimates about the proportion of people who would be convinced by the combined message to be higher than their estimates of the proportion who would be convinced by at least one of the messages. This provides support for the broad-reach-averaging model.

Message writing: Before answering the multiple choice questions, half of the participants were asked to write their own persuasive message to convince people to avoid using single-use plastic water bottles. They were then asked to categorize the message they had written as including reasons that would benefit only the receiver, benefit only others, or benefit both the receiver and others. We found that 87.4% of respondents reported writing a message that included reasons that would benefit both the message recipient and others.

Using message-writer’s own classifications, we compared the proportion of people sending a message combining both egoistic and altruistic reasons when crafting their own message relative to when people chose a message to send. We compared the group of message-writers to the other half of the participants, who had not been asked to write their own message and had only been shown the multiple choice questions. The proportion of people choosing a combined message did not differ for those who wrote their own message ($N = 253$) relative to those who chose between pre-written messages ($N = 258$), $\chi^2(1, N = 511) = 0.85$, $p = 0.36$, 95% CI $[-0.11, 0.04]$.

We also had a research professional code people’s responses. When coded by an independent researcher who was unaware of what was being tested, only 42.7% of participants explicitly mentioned benefits to both the receiver and to others, with 41.5% mentioning only benefits to the receiver and 6.3% mentioning benefits to only the receiver. It is interesting to note the discrepancy between the independent coding and how people classified their own messages. In many cases where the messages was seen by the coder as only explicitly

mentioning benefits to others, the participant who generated the message classified the reasons they gave as benefiting both the receiver and others. The types of responses people wrote largely centered around environmental benefits; it seems likely that people who use a primarily environmental argument see benefiting the environment as also providing a direct benefit to the person receiving the message, not just to others.

Table 1.5: Proportions of participants’ written messages categorized as egoistic, altruistic, or combining both reasons

	Benefit only to the receiver	Benefit only to others	Benefit to the receiver and others	Neither	Unclear
Self coding	7.5%	5.1%	83.4%	4.0%	n/a
Independent coding	6.3%	41.5%	42.7%	4.0%	5.5%

We compared the group of message-writers to the half of participants who had not been asked to write their own message. When using the independent coder classifications, the proportion of people choosing a combined message among pre-written messages (79.8%) was higher than among those who wrote their own message (42.7%), $\chi^2(1, N = 511) = 72.89$, $p < .001$, 95% CI [0.29, 0.45].

We also compared the choice of pre-written messages for those who wrote their own message to the content of their written messages, using the independent coding, and found that there was a significant difference, $\chi^2(1, N = 253) = 78.23$, $p < .001$. While 83.3% of message writers chose to send the combined pre-written message, only 42.7% had written a message independently coded as combining both egoistic and altruistic appeals.

1.8.3 Discussion

People who were motivated to convince others demonstrated a similar preference for the combined message as participants in the prior studies. This increases the generalizability of the previous results. Even in contexts when people are driven to convince others, they still demonstrate egocentric anchoring and assume that others will be similar to themselves. We

also find more evidence for the ways in which people adjust away from egocentrism. People demonstrate a “broad-reach-averaging” mental model when choosing messages, expecting that including two messages will reach a broader range of people. They still expect more people to be egoistically motivated as opposed to altruistically motivated, but the expectation of the presence of some altruistically motivated people in a population makes sending the combined message worthwhile. Additionally, we look at the messages people wrote when asked to spontaneously generate messages to persuade others. While most people classified their own messages as combining egoistic and altruistic appeals, an independent rater’s coding indicated that about the same number of people wrote combined appeals as mentioned only the altruistic benefits. The fact that they classified their own messages as benefiting the message receiver suggests that they see helping the environment as a form of benefiting the self. These results also suggest that people may perceive their own messages as incorporating more of a combination of egoistic and altruistic appeals than others perceive them to have.

1.9 Study 3b

1.9.1 *Methods*

The main goal of this study was to evaluate the accuracy of people’s predictions about message efficacy in Study 3a. Participants were recruited to this pre-registered study (<https://aspredicted.org/3bks-c46y.pdf>) using Prolific. When analyzing the results, we only included participants who responded “Yes” to the question “Have you used a single-use plastic water bottle in the past month?” in order to understand how persuasive the messages were for people who had used single-use plastic water bottles. After removing participants who did not meet the screening criteria and who failed the attention check or had the same IP address as other respondents, there were 467 valid responses. Participants were presented with either the egoistic message, altruistic message, combined message, or no message and

were asked how likely they would be to avoid single-use plastic water bottles.

1.9.2 Results

As in Study 2a, the differences between people’s behavioral intentions were small. Among the messages in, only the behavioral intentions with the largest gap, the no-message condition and egoistic-only message condition, are significantly different from one another, $t(232.4) = -3.18$, $p = 0.002$, 95% CI $[-1.06, -0.25]$.

We compared people’s predictions of the expected convincingness of each message in Study 3a to people’s behavioral intentions after seeing each of the messages. All predictions were significantly different from the actual behavioral intentions, egoistic: $M = 5.72$, $t(506) = 7.79$, $p < .001$, 95% CI $[5.63, 5.81]$, altruistic: $M = 4.74$, $t(506) = -5.15$, $p < .001$, 95% CI $[4.61, 4.87]$, combined: $M = 6.18$, $t(506) = 20.75$, $p < .001$, 95% CI $[6.09, 6.26]$. We also looked at people’s predictions about the differences in expected convincingness and compared these to the actual differences in behavioral intentions. For each pair of messages, people expected the differences in efficacy to be greater than they actually were, egoistic vs. altruistic: $M = 0.98$, $t(506) = 10.71$, $p < .001$, 95% CI $[0.85, 1.11]$, egoistic vs. combined: $M = -0.45$, $t(506) = -10.59$, $p < .001$, 95% CI $[-0.56, -0.35]$, altruistic vs. combined: $M = -1.43$, $t(506) = -18.82$, $p < .001$, 95% CI $[-1.56, -1.30]$. The motivated persuaders overestimated the differences between the convincingness of the different messages.

In addition to looking at how expected convincingness compared to behavioral intentions, we added a new way of measuring predicted behavioral intentions. We told participants in Study 3a that a separate survey had been run (Study 3b) on people who reported using a single-use plastic water bottle in the past month. We explained that each participant in that study had been shown only one message and asked them what they thought the average response would be to “How likely would you be to avoid single-use plastic water bottles?” for each group of message recipients. Before participants made their predictions,

we told them that some participants had not been shown any message, and that the average behavioral intentions for this group was 4.70. This gave participants a sense of people’s baseline behavioral intentions and was intended to help them calibrate their estimates.

Participants’ calibrated predictions were more accurate than their expected convincingness estimates. People’s estimates about the impact the egoistic-only message would have on behavioral intentions were largely accurate, $M = 5.37$, $t(501) = 0.25$, $p = 0.80$, 95% CI [5.29, 5.45]. As in the case of the expected convincingness evaluations, they overestimated the effectiveness of the combined message and underestimated the effectiveness of the altruistic message, *combined*: $M = 5.65$, $t(501) = 8.88$, $p < .001$, 95% CI [5.56, 5.73], *altruistic*: $M = 4.61$, $t(501) = -8.97$, $p < .001$, 95% CI [4.51, 4.72]. As with expected convincingness, people overestimated the differences in message efficacy, *egoistic vs. altruistic*: $M = 0.76$, $t(501) = 8.77$, $p < .001$, 95% CI [0.65, 0.86], *egoistic vs. combined*: $M = -0.28$, $t(501) = -8.99$, $p < .001$, 95% CI [-0.36, -0.20], *altruistic vs. combined*: $M = -1.04$, $t(501) = -15.32$, $p < .001$, 95% CI [-1.15, -0.93].

Table 1.6: Behavioral intentions compared to expected convincingness of messages and calibrated predictions

Message	Expected convincingness (Study 3a)		Calibrated prediction (Study 3a)		Behavioral intentions (Study 3b)	
	Mean	SD	Mean	SD	Mean	SD
Egoistic-only	5.72	1.05	5.37	0.95	5.36	1.54
Altruistic-only	4.74	1.48	4.61	1.17	5.08	1.61
Combined	6.18	0.99	5.65	0.98	5.26	1.58
No Message	n/a	n/a	n/a	n/a	4.70	1.61

1.9.3 Discussion

As in the vegetarianism studies, people’s predictions about the efficacy of the messages were largely inaccurate. Even people who are genuinely motivated to persuade others are not well calibrated in their estimates of how people will respond to the messages. While

people expected the combined message to be most convincing followed by the egoistic and altruistic single-reason messages, the differences between the behavioral intentions for these groups were not significant. After telling people the average behavioral intentions for people who did not see a message, their predictions became more accurate, particularly for the egoistic-only condition. However, they still underestimated the efficacy of the altruistic message and overestimated the efficacy of the combined message. Additionally, participants in Study 3a overestimated the differences in behavioral intentions both when looking at expected convincingness and when looking at people’s calibrated predictions, imagining that the message sent would have a greater impact than it did.

1.10 General Discussion

Over the course of five studies, we find evidence for egocentric anchoring and adjustment when people without professional training are asked to try to persuade others to engage in behaviors that can be egoistically or altruistically motivated. When choosing between an egoistic message, an altruistic message, and a message combining both types of reasons, people consistently prefer to send the combined message. Our results suggest that people start by considering what they would personally find persuasive when trying to estimate what others would find persuasive. We also find evidence for the ways in which people systematically adjust away from egocentric anchoring. They demonstrate that they consider how to reach the broadest range of people; while they recognize that not everyone will be convinced by both egoistic and altruistic reasons, including both can increase the likelihood of reaching people who will be convinced by at least one of the reasons. When evaluating how to reach as many people as possible, they operate off the assumption that people are generally more motivated by self-interested reasons than altruistic ones. We find that these results hold in the case of people who are not intrinsically motivated to persuade others and in those who are.

We find strong evidence for egocentric anchoring. While the prior literature has explored the mechanisms underlying egocentrism in contexts where people try to predict others' beliefs, our work extends this to the context of persuasion (Holmes, 1968; Karniol, 2003; Ross et al., 1977). When people evaluated the various message options, how convincing they personally found the messages consistently predicted how convincing they expected others to find them. Expected convincingness mediated the effect of personal convincingness on people's choice of the combined message. People generally expect others to be similar to them. Even in Study 3a where we explicitly selected for people with strong environmental attitudes who were sincerely motivated to persuade others to avoid using single-use plastic water bottles, they largely assumed that others would be persuaded by the messages they personally found most persuasive; they did not assume that their proclivity for environmentalism might not be widely shared.

These results align with existing work on egocentrism. People tend to overestimate the degree to which their own attributes are representative of the population (Krueger & Clement, 1994). In the context of perspective-taking, people start from their own perspective and continue to adjust if they have the opportunity to seek out information that allows them to be more accurate (Epley & Gilovich, 2006). Without clear information about the audience, people start by assuming that others are like themselves. It is worth further exploring the degree to which existing work on egocentrism characterizes people's behavior and message choices in contexts where the aim is to shift other people's behavior.

Despite egocentrism's important role, it does not fully characterize people's expectations about efficacy; there is not a perfect match between people's personal views and their predictions about others. We explore the types of adjustments that people make, over and above anchoring egocentrically. We found evidence for multiple different mechanisms driving the "adjustment" portion of "egocentric anchoring and adjustment". While egocentrism largely predicted people's behavior, people deviated from egocentrism in systematic ways. One of

the ways in which people did not demonstrate egocentrism was in their estimates of how convincing people would find the egoistic message relative to the altruistic one. Across all three domains, people expected the egoistic message to be more convincing than the altruistic message, even after controlling for how personally convincing they would find each message. This aligned with research on the norm of self-interest; when trying to convince others, message senders assumed that the message recipients would be largely persuaded by how the change in behavior would benefit themselves. In the absence of concrete information about their audience, people expect others to be more self-interested than altruistic.

We did not find evidence that people who had a strong moral stance on the behavior they were promoting would only want to use moral reasons. Even in the case where we selected for people who were morally aligned with the subject, people who had stronger moral positions were consistently more likely to include the egoistic reason than those who were not. We did not find evidence for “signaling” one’s altruism by only sending the altruistic message; regardless of their personal opinions, few people sent the altruistic-only message across studies. We were able to rule out theories that suggest that those with strong moral beliefs would find it taboo to mix moral and non-moral reasons or that they care about whether people engage in the behavior for the “right” reasons. This could possibly be due to the nature of the topics we selected; it is plausible that people would be more concerned about moral taboos and the sincerity of people’s motivations on issues that are more universally morally condemned than not wearing masks, eating meat, or using plastic. Perhaps the morally motivated persuaders were more likely to include the egoistic reason because they were willing to use whatever strategies necessary to shift people’s behavior and believed this would be the most effective strategy.

The other way in which people deviated from egocentrism was through their use of the broad-reach-averaging approach to persuasion. Despite people’s expectation that self-interested reasons will persuade others more than altruistic reasons, they still choose to

send the message combining both reasons. In Study 3a, we found that people expected a larger proportion of an imaginary sample population to be persuaded by at least one of the single-reason messages than by the combined message. This suggests that people chose the combined message not because they expected the two reasons to be more persuasive together, but because they expected including the two reasons to reach a broader range of people.

This differs from the model proposed in work on the presenter’s paradox in that it allows for consideration that different people may find different reasons persuasive. In the presenter’s paradox, the assumption is that all components making up a bundle will be received positively. Presenters therefore demonstrate an additive approach, assuming that adding more positive components to a bundle will be received better than fewer components. However, this logic does not apply in the case of behaviors that can be motivated for egoistic or altruistic reasons. People intuit that others will differ in terms of how well they receive each reason. We find that motivated persuaders expect 35% of a random sample of the population to be convinced by only the egoistic message and 19% to be convinced by only the altruistic message. They expect only 36% to find either reason convincing. An averaging approach on the level of the individual would predict that people would only send the single reason they expected to find most effective, which is not what we find. Instead, we clearly see that people send both egoistic and altruistic reasons in order to reach as many people as possible.

Our results from study 2a indicate that this preference is likely driven primarily by the number of reasons, not the mixing of reasons of different types. When given the choice of two-reason messages over single-reason messages, people prefer to send the two-reason messages, even when these messages do not combine egoistic and altruistic reasons. While people begin with egocentrism, they do not solely assume that people will be like them; they adjust their message-sending strategy to account for a diversity of perspectives. These findings suggest that people perform their own pared-down, intuitive version of Total Unduplicated Reach and Frequency analysis when choosing between pre-written messages that contain

relatively small numbers of reasons for engaging in a new behavior.

Overall, we find evidence that laypeople use a sophisticated, nuanced strategy when trying to persuade others. While people do assume that others will be convinced by the messages they find most convincing, they adjust their assumptions to account for a diversity of perspectives across the population. When making inferences about the population, people assume the norm of self-interest to an extent; they do not expect people to be solely motivated by self-interest, but they do expect this motivation to take precedence for most people over the desire to help others.

We find that these results hold both in the case of people without professional persuasive training who are asked to merely imagine that they are trying to convince others and in the case of people who are sincerely motivated to persuade others. While many people in the mask-wearing study were in support of masking at the time the studies were run, most of the people in the vegetarianism study were not vegetarians. However, in both of these cases and in the plastic bottle study where we selected people who were interested in convincing others, people demonstrated strong egocentric anchoring and adjusted away from egocentrism in similar ways.

We also looked at the types of messages motivated persuaders wrote when asked to write their own messages. When they classified their own messages, the vast majority (87%) reported writing a message that would benefit the receiver and others. Interestingly, when coded by an independent rater, only 42% of the messages contained explicit mentions of benefits to self and others, with 40% only mentioning benefits to others. This discrepancy was possibly due to the fact that message writers saw their emphasis on the environmental benefits of avoiding single-use plastic water bottles as also being beneficial to the self. While some people explicitly mentioned how protecting the planet benefits the receiver, many people simply mentioned the environmental benefits without further explanation. Based on the messages people wrote about the ways in which avoiding disposable plastic water bottles

conserves natural resources, reduces plastic waste, protects the oceans and wildlife, etc., it seems plausible that they are assuming that others will see the self-benefiting value in these reasons. Yet, when given the opportunity to include explicitly self-oriented benefits in the pre-written messages, these same participants largely choose the combined message (80%). They see the value in explicitly mentioning the egoistic reason when presented with it, but they do not spontaneously generate this reason themselves.

Participants followed a similar pattern as was seen in prior work on moral reframing (Feinberg & Willer, 2015, 2019). Liberals and conservatives write arguments focused on their own moral convictions; however, when given a choice of pre-written messages, they choose the morally reframed message that aligns with the values of their target audience. When generating messages, people spontaneously seem to demonstrate an even stronger pattern of egocentric anchoring than when choosing between messages that have already been written. When presented with other options, people adjust away from their egocentric perspectives and choose what to say in a more strategic way. The adjustment strategies that people use when choosing between messages, including a range of reasons to reach a broad audience and assuming the norm of self-interest, show up less strongly when people are asked to write messages themselves. It is worth better understanding the differences between the strategies people use when choosing between messages and writing their own messages to try to convince others to change their behavior.

In addition to looking at people's expectations about which messages would be convincing, we looked at whether these predictions were accurate. In the case of vegetarianism and when encouraging people to avoid disposable water bottles, people's predictions were largely inaccurate. The differences between people's behavioral intentions in both Study 2b and Study 3b were relatively small compared to what had been predicted. In Study 2b, participants expected others to be most convinced by the message encouraging people to go vegetarian for their health and to save money, the message that combined two egoistic reasons. They

expected the altruistic messages to be the least effective, whether they included two altruistic reasons or only one. In reality, the message talking about both helping animals and helping the environment yielded one of the highest behavioral intentions, and the message with two egoistic reasons yielded one of the lowest behavioral intentions. In Study 3b, people expected the messages containing the egoistic reason, either in conjunction with the altruistic reason or on its own, to be significantly more effective at persuading others than the altruistic-only message. When looking at behavioral intentions, there were no significant differences in peoples' behavioral intentions across the three messages. Message senders' expectations about the convincingness of the egoistic reasons align with work on "extrinsic incentives bias" (Heath, 1999). While this bias was studied in the context of managers overseeing employees, our results follow the same pattern of people overestimating the extent to which people care about extrinsic incentives relative to intrinsic ones.

The discrepancy between people's expectations and behavioral intentions decreased after providing participants with additional information to calibrate their estimates. In addition to measuring expected convincingness, we asked participants to predict behavioral intentions after telling them the baseline behavioral intention for plastic bottle users who had not been shown any message. In this case, people's predictions regarding the egoistic-only message were accurate, and their estimates regarding the altruistic and combined messages were more accurate, though still significantly different from the message receivers' behavioral intentions. This suggests that people are able to adjust their predictions when provided with new information, including in ways that counteract the "extrinsic incentives bias" when considering the egoistic-only message.

These findings offer new insights into the strategies laypeople use when trying to persuade others. Given the extent to which people are influenced by those they know, it is worth better understanding the mechanisms driving the choices of those trying to influence others. While non-professionals' approaches to influence others' behavior follow clear patterns of

egocentric anchoring and adjustment, there is little evidence that their approaches are effective. This topic merits further exploration in more naturalistic settings through observing real conversations and measuring subsequent behavior change. It could also be useful to better understand the degree to which egoistic, altruistic, and combined messages shift behavior. The results from the existing literature conflict with our results and with each other (Feiler et al., 2012; Kareklas et al., 2014). The extent to which egoistic and altruistic messages are likely to shift behavior is likely context-dependent, and it would be valuable to better understand when these reason types are more or less likely to be effective.

1.11 Conclusion

We contribute to a growing body of literature on how people go about persuading others. When choosing between egoistic, altruistic, and combined messages, people prefer to send messages combining egoistic and altruistic reasons rather than sending either type of message on its own. This preference for two-reason messages persists when including two reasons of the same type, either both altruistic or both egoistic. People’s choices are driven by egocentric anchoring; persuaders’ message choices are primarily driven by what they would find personally convincing. They also demonstrate systematic adjustments away from pure egocentrism, sending two-reason messages to reach as many people as possible with at least one of the reasons. Their assumptions about the population are influenced by their assumption that others will be more motivated by self-interest than altruism.

This new understanding illuminates the mechanism underlying social interactions that many of us encounter on a regular basis. Understanding how laypeople persuade others could allow public advocacy campaigns, which rely on word-of-mouth, to be designed more effectively. This is particularly useful in the context of behaviors that are both good for others and good for the person engaging in the behavior. Behaviors that naturally provide a dual benefit to the self and others can be performed at a lower cost to the actor; they

reduce or entirely eliminate the undesirable degree of self-sacrifice required by many altruistic acts. While not everyone wears face masks in relevant contexts or is mindful of their meat and plastic consumption, surely more people do so than if these behaviors were only seen as benefiting others. It is worth better understanding how these widespread social dynamics play out to advance the theory and practice of persuasion.

CHAPTER 2

THE POWER OF CONSISTENCY: THE EFFECTS OF WORD-DEED ALIGNMENT ON BEHAVIORAL INTENTIONS

2.1 Abstract

While advocates often model the behaviors they are promoting to others, there are also times where their words and actions do not line up. In this paper, we seek to understand the ways in which an advocate's word-deed alignment and misalignment impact people's intentions to adopt the behaviors an advocate is promoting. We explore this in the case of behaviors that have not yet gained widespread acceptance and that can be done for self-interested or other-oriented reasons. In study 1, we cross a solar panel ambassador's behavioral consistency (owning or not owning solar panels) with the appeal type used: either egoistic, altruistic, or both. In study 2, we use a similar study design focusing on vegetarianism, varying whether a pro-vegetarianism advocate is seen eating a vegetarian meal or a meal containing meat. In both studies, we find that inconsistency increased dislike of the ambassador, and the extent to which people dislike the ambassador plays a greater role in determining people's behavioral intentions than their inferences about the ambassador's beliefs and their perceptions of the ambassador's judgmental attitudes and persuasive intentions towards them. When their inferences about the ambassador's beliefs do play a role, the extent to which the ambassador believes that the behavior will be beneficial to the self has a larger impact on behavioral intentions than people's second-order beliefs about the benefits to others or the degree to which a behavior aligns with social norms. In study 3, we evaluate the robustness of the previous results, continuing to vary the consistency of the advocate's behavior while giving participants a verbal indication of the ambassador's true beliefs about solar panels. Participants imagined overhearing the ambassador having a conversation about solar panels with his wife, in which

he spoke positively about the benefits of installing solar panels. As in the previous studies, disliking the ambassador based on the inconsistency was the primary predictor of people's intentions to install solar panels. The impact of the advocate's behavior on intentions was not attenuated by the ambassador's private statement of beliefs; participants continue to dislike the ambassador who behaves inconsistently and do not give him the benefit of the doubt. These studies offer a more nuanced perspective on the psychological mechanisms determining people's behavioral intentions in response to word-deed misalignment, showing that people are more likely to have a direct negative reaction to inconsistency than they are to make careful inferences about beliefs from behavior.

2.2 Introduction

Imagine you are walking through your town and are approached by a man selling solar panels. He tells you about the benefits of solar panels, both for the environment and to you financially. Later that day, you happen to drive by the solar panel ambassador sitting on what looks to be his own front porch. As you look at the roof of his house, you notice that, despite being in a sunny location, there are no solar panels installed.

There are many ways one might react in this situation. People might have a visceral aversive reaction toward the ambassador, leading them to not give the prospect of installing solar panels from him a second thought. Alternatively, they might wonder why this discrepancy occurred and consider possible explanations. It could be that the ambassador earns a commission on sales and was trying to persuade you, despite not believing what he said. Perhaps the solar panels are not a good financial investment, or maybe they do not actually benefit the environment. There is also a chance that the things he said were true and there is some other explanation for why he has not installed them on his own home yet. In any case, it is likely that one would be less inclined to purchase panels from this man after observing inconsistency, compared to if they had seen panels perched atop his roof.

Most people value behaving consistently, both because of how it impacts their own self-perception and how others perceive them (Cialdini & Goldstein, 2004). Despite the influential power of “practicing what you preach”, people in persuasive roles often behave hypocritically (Han et al., 2017; Lockwood & Kunda, 2000). Deviations from this norm can be striking, particularly if there is not an obvious explanation available. Without knowing why someone’s words are not aligned with their actions, people are less likely to adopt a behavior after seeing that someone promoting it is not engaging in it themselves (Kraft-Todd et al., 2018).

There are several possible reasons why individuals may be reluctant to engage in a behavior after observing that a vocal advocate does not practice it. We investigate people’s interpretations of an advocate’s beliefs, their own attitudes toward the advocate, and their perceptions of how the advocate views them. People may make inferences about an advocate’s beliefs, considering several reasons why someone may not behave consistently (Graham et al., 2015). They might infer that the advocate is being paid to say things that contradict their true beliefs, or they may charitably consider that there is some other reason why it is challenging to follow through; perhaps a solar panel ambassador lives in a rented house and does not have the right to install panels. Alternatively, people may simply not want to purchase a product or adopt a behavior being promoted by someone they do not like or trust, due to inconsistency. Lastly, people may adopt the behavior out of concern for being judged by a consistent advocate, or, in the case of encountering an inconsistent ambassador, they may choose to actively resist the persuasion attempt and avoid the behavior as a means of asserting autonomy.

In this paper, we explore the psychological mechanisms underlying people’s decisions to engage in a behavior after observing an advocate whose actions either align or do not align with their words. What is the process by which the consistency of an advocate’s words and actions affects someone’s behavioral intentions?

To increase generalizability, we focus on behaviors that can be promoted for either self-

interested or other-oriented reasons. It is not clear how someone’s words and actions influence others when there are several plausible reasons for engaging in a behavior. If someone promoting solar panels highlights the ways in which solar panels help the environment and save money but does not have panels on their own house, people could interpret the lack of consistency in multiple possible ways. They may assume that the panels are not good for the environment or that they are a poor financial choice. Depending on whether they care more about saving money or going green, they may dislike the ambassador for recommending solar panels that would not align with their values. By including behaviors that can be motivated for selfish or altruistic reasons, we can gain a more robust understanding of the complexity of people’s reactions to behavioral inconsistency.

It is useful to understand people’s reactions to these types of behaviors because there is typically a much lower cost to engage in them than in behaviors that are purely altruistic. Taking an action that is good for the self and good for others does not involve self sacrifice. When encouraging positive social change through individual behavior change, it is potentially more effective to focus on behaviors that are likely to elicit less resistance by aligning both selfish and altruistic motivations.

In this paper, we seek to better understand how an advocate’s behavior influences people’s actions in the case of behaviors that can be construed as beneficial for self-interested or other-oriented reasons. We focus on people’s intentions to adopt these behaviors. In the following section, we explore various theories that offer frameworks for answering this question. We consider people’s inferences about an advocate’s beliefs, their personal attitudes toward the advocate, and their perceptions of the advocate’s attitudes toward them.

2.2.1 Credibility-enhancing displays

One possible explanation is that consumers rely on ambassadors’ actions as an indication of their true beliefs. Credibility-enhancing displays are actions that provide support for an

actor’s verbally-expressed beliefs. They are costly, meaning that someone would be unlikely to engage in them if their beliefs did not genuinely match their words. Cultural evolutionary theory posits that CREDs evolved to solve a common problem: that people can easily mislead others by claiming to believe certain things without offering evidence. CREDs can be powerful indicators of people’s true beliefs, regardless of whether a behavior is primarily motivated by self-interest or altruism (Henrich, 2009; Henrich & Gil-White, 2001).

The inferences people make become more complex when a behavior can be promoted for either self-interested or other-oriented reasons. Kraft-Todd et al. (2018) posit that CREDs explain why people are more likely to install solar panels when people see that the solar panel ambassador has panels on their own home. They measure people’s second-order beliefs: their beliefs about the ambassador’s beliefs. However, the ambassador in their study mentions both the altruistic and selfish benefits of solar panels. In their explanation for how people’s second-order beliefs influence behavior, they do not differentiate between the inferences people might make about the ambassador’s beliefs about the financial or environmental benefits of installing solar panels. It is not clear how someone’s words and actions influence others when there are many plausible reasons for engaging in a behavior and deception is potentially at play.

In instances where a behavior benefits the individual performing it, CREDs suggest that the person believes the behavior to be in their best interest (Henrich, 2009). One classic example centers around evaluating how safe it is to eat a mushroom. It is easy for someone to say that a mushroom is safe to eat. However, their claim is much more credible if they eat it themselves while their target is watching (Kraft-Todd et al., 2018). Eating the mushroom serves as an effective CRED because it is consistent with a belief that the mushroom is safe, and someone would be unlikely to do it if their genuine belief was that the mushroom was dangerous. In contrast, saying the mushroom is safe but not eating it allows for many alternative interpretations. The person making the statement may sincerely believe that the

mushroom is safe despite not wanting to eat it at that moment, or they may be deceiving the listener and the mushroom may not be safe to eat. CREDs make the behavior easier to interpret by demonstrating that the person’s stated beliefs are consistent with their behavior.

In altruistic contexts, CREDs show the extent to which someone believes in the behavior they are promoting (Henrich, 2009). Engaging in low-cost activities like discussing a charity or posting about it on social media may not demonstrate one’s true commitment to a cause. Alternatively, making a donation can serve as a powerful CRED, signaling to others that the individual is committed to this action. This type of signaling plays an important role in group cultural learning (Henrich, 2009). People are more likely to engage in behaviors that are personally costly but benefit the group when they can see that others are genuinely engaged in those behaviors as well.

2.2.2 Other theories

In addition to focusing on how an advocate’s behavior affects inferences about the advocate’s beliefs, it is worth considering other theories that might explain people’s behavior in similar scenarios. The social dynamic between an individual and an advocate might impact someone’s willingness to adopt a behavior, above and beyond their inferences about an advocate’s beliefs. We consider the impact of people’s attitudes toward an advocate and their perceptions of attitudes from the advocate.

Attitudes toward the ambassador

People’s attitudes toward an ambassador promoting a new behavior may influence their likelihood of adopting that behavior, independent of their perceptions of the advocate’s beliefs. In the case of an advocate behaving consistently with the message they are promoting, such as when a solar panel ambassador has solar panels on their own house, people’s positive attitudes toward the ambassador might motivate them to purchase panels themselves.

Positive motivators When people see an advocate behave consistently, they may admire the advocate and feel inspired to follow their lead. Talking to someone encouraging you to engage in a new behavior that could positively impact you or others could be inspiring if the behavior is seen as relevant and attainable (Lockwood & Kunda, 1997, 2000). Role models can be inspiring both when a behavior is seen as primarily benefiting the self or benefiting others. While inspiration can work both in the cases of egoistic and altruistic behaviors, there is evidence suggesting that people may be more inspired by someone behaving for moral reasons than for self-interested ones (Bolderdijk et al., 2018). Morally motivated advocates might be more likely to elicit inspiration than those highlighting self-interest, as long as their behavior does not threaten the observer’s moral self-concept (Minson & Monin, 2012).

Negative motivators While aligned behavior could elicit admiration, inconsistency between an ambassador’s words and deeds can lead people to dislike the ambassador. This can happen regardless of whether the behavior is being promoted for egoistic or altruistic reasons. Whether the behavior is seen as primarily moral or amoral will impact the possible mechanism by which disliking the ambassador has an effect on people’s behavior.

Moral hypocrisy When an action is seen as predominantly moral, promoting it verbally while behaving differently may be seen as hypocritical. Word-deed misalignment is seen as hypocrisy when there is a moral valence, particularly when it appears as though someone received a moral benefit that they did not deserve (Effron et al., 2018). There are several reasons why someone may behave hypocritically that are not always apparent from just observing behavior. Someone may be trying to seem more moral than they are. They might be trying to motivate others by appealing to others’ moral values, even if they do not share those values themselves. They may also simply find it hard to follow through on a moral action, even if they genuinely endorse it. People judge others differently depending on the explanations they have access to for why someone behaved hypocritically. They evaluate

false signaling more negatively than committing a personal moral failing (Jordan & Sommers, 2020). People’s attributions around why someone acted hypocritically determine the degree to which they dislike someone.

Seeing an advocate of a behavior as hypocritical may make people less likely to adopt the behavior being promoted. While it is clear that people do not like hypocrites, it is less clear why this impacts people’s willingness to adopt a behavior. One theory posits that seeing hypocrisy leads to a sense of distrust. Feeling distrust toward others makes people feel like others are trying to exploit them, affecting people’s personal moral standards and making them more likely to behave hypocritically themselves (Weiss et al., 2018). This suggests that someone may not adopt a behavior after seeing an advocate behave hypocritically due to distrust leading them to behave in a more self-serving way, even if they would otherwise feel morally aligned with that behavior.

At face value, the mechanism driving the effect of witnessing hypocritical behavior on behavioral intentions looks similar to the mechanism driving credibility-enhancing displays. In both cases, people are looking to an advocate’s actions to understand how committed that individual is to a certain moral belief. The main difference between these two theoretical approaches lies in what they emphasize. In the case of credibility-enhancing displays, the focus is on inferring belief from an advocate’s consistent behavior. Theories related to moral hypocrisy stress inferring a lack of belief from inconsistent behavior and instead forming inferences about the person’s moral character. Hypocrisy centers around the negative effects of behaving inconsistently more so than the positive benefits of behaving consistently.

Amoral dislike When an action is seen as primarily benefiting the self, people may dislike someone promoting it if they behave inconsistently, even when the action has no moral valence. This can damage an advocate’s credibility, making others less inclined to adopt the behavior. Work on source credibility shows that trusting a source has a positive persuasive impact (Wiener & Mowen, 1986). A communicator’s character significantly affects the persuasiveness

of their message (Wilson & Sherrell, 1993). Additionally, work on celebrity spokespersons shows that peoples' attitudes toward a celebrity can transfer to the endorsed brand, for better or worse (Min et al., 2019). People look to adopt the behavior of people who they trust.

The impact of an advocate's inconsistency may go above and beyond people's perceptions of the advocate's beliefs about the behavior. While credibility-enhancing displays focus on the advocate's behavior as an accurate indicator of beliefs, this work suggests that people's dislike of the ambassador might affect their behavioral intentions separate from their inferences about the ambassador's beliefs. People may be less likely to purchase a product that would improve their life if they do not like the celebrity endorsing it (Min et al., 2019). Similarly, even if people believe a behavior would benefit themselves, their negative attitudes toward the ambassador may spillover into their behavioral intentions.

Attitudes from the ambassador

While we primarily focus on CREDs and people's attitudes toward the ambassador when considering the influence of an advocate's behavioral consistency on people's subsequent behavior, we also wanted to acknowledge that perceptions of attitudes from the ambassador could contribute as well. In particular, people's sense of being judged by the ambassador and their awareness of the ambassador's desire to persuade them could influence their likelihood of adopting the behavior being promoted.

When an advocate's words and actions match, the perception of being judged by that advocate may influence the listener's behavior, particularly when the advocate is a member of the listener's community. Many people care deeply about being seen as moral and will behave in ways that are more likely to lead others to see them that way (Aquino & Reed II, 2002; Minson & Monin, 2012; Monin & Jordan, 2009). Taken a step further, the desire to feel a sense of belonging with a social group and, conversely, the fear of being excluded from that group are powerful motivators for people to conform (Baumeister & Tice, 1990). In our

studies, we specify that the advocate is a member of the participant’s community. To the extent that someone sees the advocate as an integral member of the community, a fear of judgement for not adopting the behavior being promoted could strongly drive behavior.

When an advocate’s words and actions do not match, people would likely be less concerned about the advocate’s judgment toward them. Instead, they may see the ambassador’s inconsistency as a sign that the ambassador may have ulterior motives in trying to persuade them. When people become aware of a persuasion attempt, they shift their attention away from the content of a message toward protecting themselves from social pressure (Friestad & Wright, 1994). This perceived threat can trigger psychological reactance, a desire to reclaim one’s autonomy and freedom (Brehm & Sensenig, 1966; Feiler et al., 2012). This could take the form of being even less likely to engage in the behavior being promoted than if one had not encountered the ambassador at all.

2.2.3 Roadmap

In the following studies, we aim to deepen our understanding of the psychological mechanisms underlying people’s decision to adopt a new behavior when encountering an advocate who either practices or fails to practice the behavior they are promoting. We investigate the relative effects of people’s inferences about the ambassador’s true beliefs compared to their personal attitudes towards the ambassador in their behavioral intentions. Our study designs are adapted from the scenarios used in Kraft-Todd et al. (2018). Across all three studies, we find that disliking the ambassador is the primary mediator of people’s intentions to adopt a new behavior after observing an inconsistent (vs. consistent) ambassador. Even when splitting by the appeal type used, second-order beliefs play a secondary role; people’s personal attitudes toward the ambassador serve as the main determinant of their intended behavior when using egoistic, altruistic, and combined appeals.

In study 1, we investigate people’s responses to a solar panel ambassador who either does

or does not have panels on his own home. We build on Kraft-Todd et al.'s scenario, adding a neutral control condition in which the ambassador's solar panel ownership is not specified. We also separate out the type of appeal the ambassador uses; rather than always using a combined egoistic and altruistic appeal, we have egoistic, altruistic, and combined conditions. In addition to the second-order beliefs scale, we add measures to test other psychological mechanisms. We find that people's dislike for the ambassador has a greater negative effect on their behavioral intentions than their inferences about whether the ambassador believes the claims he is making about solar panels. In further support of this narrative, we find that the magnitude of the negative effect of the ambassador's behavioral inconsistency is larger than the positive effect of consistency.

In study 2, we extend these findings to the domain of vegetarianism. In the scenario, people imagine encountering a vegetarian ambassador who is later seen eating a vegetarian meal or a meal containing meat. Again, we find that people's personal negative attitudes towards the inconsistent ambassador predict their behavioral intentions more strongly than their views about the ambassador's beliefs. The negative impact on behavioral intentions is greater in response to behavioral inconsistency than the positive effect of seeing the advocate's actions match their words.

In study 3, we test the robustness of the prior findings using a manipulation of the ambassador's true beliefs. We use a scenario adapted from Kraft-Todd et al., asking participants to imagine that they overhear the ambassador on the phone with his wife, privately stating his enthusiasm for solar panels. We manipulate the ambassador's behavioral consistency as in the previous studies, building on Kraft-Todd et al. by including a control condition where the behavior is not specified, having separate conditions in which we vary the appeal type used, and adding measures of people's attitudes toward the ambassador and perceptions of the ambassador's attitudes toward them. If people's inferences about the ambassador's beliefs were the primary driver of their behavior, we would expect to see the difference in behavioral

intentions between the consistent and inconsistent conditions reduced. Instead, we found that behavioral inconsistency still has a negative effect on people’s behavior, due to people disliking the ambassador. Actions impact people’s behavioral intentions more than words.

2.3 Study 1

2.3.1 *Methods*

This study aimed to investigate the psychological mechanisms underlying people’s willingness to adopt a new behavior after encountering an advocate who either does or does not engage in that behavior themselves. More specifically, we look at the impact of the presence or absence of solar panels on the home of a solar panel ambassador on people’s intentions to purchase solar panels, modeled after a scenario used in Kraft-Todd et al. (2018). We build upon Kraft-Todd’s scenario, adding a neutral control condition in which it was not specified whether the solar panel ambassador had panels on his home. The presence of this control condition helps to differentiate between the unique psychological mechanisms impacting decisions after seeing consistent behavior compared to those impacting decisions when seeing inconsistent behavior. This allows us to understand whether the impact of behavioral consistency is primarily due to the benefits of consistency as opposed to the cost of inconsistency.

Participants were asked to imagine the following, based on a scenario adapted from Kraft-Todd et al. (2018).

“Imagine that the town you live in is taking part in a campaign to install more solar panels in residential homes in partnership with a program called Solarize. To spread the word about the program, Solarize recruits local community leaders to volunteer their time working on the campaign, known as solar ambassadors.”

On the following page, they were shown a description of Jonathan, the local solar panel ambassador. When reading that description, participants in the treatment conditions were

shown one of the following:

Consistent condition: “He also decided to install solar panels on his own home through the Solarize program.”

Inconsistent condition: “He does not have solar panels on his home and decided not to install solar panels on his own home through the Solarize program.”

In addition to the primary manipulation of the solar panel ambassador’s behavior, we also manipulated the type of persuasive appeal used: egoistic, altruistic, or a combination of the two. This serves as a robustness check to verify whether the psychological mechanisms at play are primarily relevant in egoistically or altruistically motivated behaviors. Additionally, it allows us to see whether people primarily focus on the egoistic or altruistic reason when both reasons are used. In the aforementioned neutral control condition, the reason why participants should consider installing solar panels was not specified. Overall, the study employed a $2 \times 3 + 1$ between subjects design, varying the presence or absence of solar panels on the ambassador’s home and the type of persuasive appeal (egoistic, altruistic, or both), with the addition of a neutral control condition in which it was not specified whether the ambassador had panels on his home and no specific reasons were given for why participants should consider installing solar panels.

The reason given was manipulated in the following way. When explaining to the participants why they should consider installing solar panels, they were told by the ambassador:

Egoistic appeal: “Residential solar panels can help you save money on energy bills.”

Altruistic appeal: “Residential solar panels can help conserve the environment.”

Egoistic & altruistic appeal (order of reasons counterbalanced): “Residential solar panels can help you save money on energy bills and conserve the environment.”

Control condition: “Consider taking advantage of the benefits of the Solarize program!”

After viewing the message from the ambassador, participants were asked to rate their likelihood of installing solar panels, both through Solarize and in general, using a 7-point Likert scale. They were then asked a series of questions about the potential mediators, including their second-order beliefs about the ambassador, their personal attitudes toward the ambassador, and their expectations about how the ambassador would perceive them. Additionally, they were asked about their attitudes toward Solarize and toward solar panels in general.

Participants were recruited using Amazon Mechanical Turk for this pre-registered study (https://aspredicted.org/GDY_8K2). After removing participants who failed the attention check or had a duplicate IP address, there were 586 valid responses remaining from the 593 respondents who completed the survey.

2.3.2 Results

Dimensionality reduction

To gain insight into how the ambassador’s behavioral consistency impacts behavioral intentions, we examined several factors that may play a role. These included the second-order beliefs scale from Kraft-Todd et al. (2018), which assesses participants’ perceptions of the ambassador’s beliefs about solar panels. Additionally, we measured participants’ attitudes toward the ambassador and their perception of the ambassador’s intentions and attitudes toward them. We used a question about the “judgment gap” between the ambassador and listener adapted from Hoogendoorn et al. (2019).

We ran a principal component analysis with varimax rotation on the second-order beliefs scale. In Kraft-Todd et al. (2018), they aggregated all measures of this scale. However, after

running a PCA, we identified three distinct components (Table 2.1). This allowed us to get a more detailed understanding of which of these variables were driving people’s behavioral intentions. We removed one question from the original scale that significantly loaded onto two of the components: “How much does Jonathan believe that purchasing solar panels through the Solarize program is going to help make people happier or healthier?”

Table 2.1: Solar panel second-order beliefs

Component	Questions making up this component
	All questions begin: “How much does Jonathan believe that purchasing solar panels through the Solarize program...”
Social norms	... is something that many people in your community would do if given the opportunity? ... makes someone a good person in the eyes of people in your community? ... is what people in your community think everyone should be doing? ... helps a person’s reputation in the eyes of people in your community?
Egoistic: Saving money	... is going to help you to save more money in the long run? ... is going to lead to you having lower bills (e.g. for utilities)? ... will eventually pay itself off (i.e. the benefit will at least match the cost) at some point in the future? ... will result in you spending less money and so cut your costs?
Altruistic: Helping the environment	... support protecting the environment or conserve natural resources? ... is going to make life better for future generations of people? ... will do some part to contribute to the greater good? ... is going to help make people happier or healthier?

We also ran a PCA on the questions related to people’s attitudes toward and from the ambassador, which yielded 4 separate components (Table 2.2). The first component aggregates people’s attitudes regarding how much they like or dislike Jonathan. While we had originally separated out positive motivators like admiration from negative motivators like moral hypocrisy and amoral dislike, the PCA results suggested that the study participants were thinking about these questions in similar ways. We used the component formed from

questions about Jonathan’s intentions to change your values and respect for your values as proxies for persuasion awareness. Interestingly, we found that people’s estimations of Jonathan’s judgment toward them formed a separate component from people’s predictions about how Jonathan would judge their morality. It is possible that people interpreted the former as including both moral and amoral judgment, such as judgment about making a poor financial decision, whereas the latter only emphasized morality.

Table 2.2: Solar panel attitudes toward and from the ambassador

Theoretical constructs	Component	Question
Moral hypocrisy; Amoral dislike; Admiration	Dislike	How hypocritical is Jonathan? How annoyed would you be after talking to Jonathan? How much do you admire Jonathan? (reverse-coded)
Persuasion awareness	Change your values	How likely is it that Jonathan actively wants to change your values around helping the environment? How likely is it that Jonathan respects your values around helping the environment? (reverse-coded)
Non-specific judgment from ambassador	Judge	If you do not install solar panels, how likely is it that Jonathan will judge you negatively?
Moral judgment from ambassador	Judgment gap	If Jonathan found out you did not install solar panels, he would think you were: 1-extremely immoral, 7-extremely moral (reverse-coded)

Identifying possible mediators

We used multiple regression to identify possible mediators of the effect of the ambassador’s behavioral consistency on participants’ behavioral intentions. We regressed people’s intentions to purchase solar panels through Solarize on the consistency of the advocate’s behavior, their

second-order beliefs about the ambassador, their attitudes toward the ambassador, and their perceptions of how the ambassador was relating to them. Participants’ behavioral intentions were measured by asking them to evaluate their likelihood of purchasing solar panels through the Solarize program using a 7-point Likert scale in response to the following question: “How likely would you be to purchase solar panels through the Solarize program?” For second-order beliefs, we began by using a measure that aggregated all the questions related to second-order beliefs rather than using the components we isolated using PCA. Aggregating second-order beliefs in this way allowed us to compare the results to Kraft-Todd et al. (2018) more directly, as this is how they analyzed their data.

When controlling for aggregated second-order beliefs and the ambassador attitude components, the effect of the ambassador’s consistency on behavioral intentions was no longer significant, $\beta = -0.11$, $t(586) = -0.69$, $p = 0.49$. This suggests that the other variables in the regression could potentially serve as mediators. As in Kraft-Todd et al. (2018), the aggregated second-order beliefs had a significant positive effect on people’s intentions to install solar panels through the Solarize program, controlling for ambassador attitude components $\beta = 0.21$, $t(586) = 3.52$, $p < .001$. However, the results indicate that second-order beliefs do not entirely explain the results. People’s dislike toward the ambassador has a larger effect on behavioral intentions; when people dislike the ambassador, they are less likely to report a willingness to engage in the behavior, controlling for second-order beliefs and the other ambassador attitude components, $\beta = -0.57$, $t(586) = -10.17$, $p < .001$.

We ran a similar regression using the separate second-order belief components identified using PCA (Table 2.3). This allowed us to gain a better understanding of the relative importance of people’s views on the ambassador’s beliefs about how much solar panels save money, help the environment, and align with social norms. People’s expectations about the egoistic, money-saving benefits of solar panels had the largest effect of the three components, $\beta = 0.13$, $t(585) = 1.90$, $p = 0.058$. When separating out various aspects of second-order

beliefs, people’s behavioral intentions were primarily driven by the extent to which they thought the ambassador thought solar panels would help them save money. Disliking the ambassador had a larger impact on people’s behavior than any of the second-order belief components, $\beta = -0.54$, $t(585) = -9.32$, $p < 0.001$.

Table 2.3: Disliking the ambassador predicted people’s behavioral intentions more so than second-order beliefs. Among second-order beliefs, egoistic second-order beliefs had the greatest impact.

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	4.95***	0.54	3.89	–	6.01	<0.001
Consistent	–0.15	0.16	–0.47	–	0.17	0.362
Social Norms SOB	0.06	0.06	–0.06	–	0.18	0.304
Egoistic SOB	0.13	0.07	–0.00	–	0.27	0.058
Altruistic SOB	0.07	0.08	–0.08	–	0.22	0.360
Dislike	–0.54***	0.06	–0.66	–	–0.43	<0.001
Change Values	0.02	0.07	–0.12	–	0.16	0.791
Judge	0.07	0.04	–0.02	–	0.15	0.108
Judge Gap	–0.08	0.07	–0.22	–	0.07	0.298
Observations	594					
<i>R</i> ² / <i>R</i> ² adjusted	0.360 / 0.352					
		* <i>p</i> < 0.05	** <i>p</i> < 0.01	*** <i>p</i> < 0.001		

Mediation model

The prior results suggested that egoistic second-order beliefs and disliking the ambassador were the two primary candidates as mediators of the effect of the ambassador’s consistency on participants’ behavioral intentions. We ran a parallel mediation model including these two variables. Together, egoistic second-order beliefs and disliking the ambassador fully mediate the effect of consistency on behavioral intentions, total effect: 1.09, 95% CI [0.83, 1.35], $p < 0.001$; direct effect: 0.08, 95% CI [–0.18, 0.33], $p = 0.55$.

The indirect effect of disliking the ambassador is larger than the effect of egoistic second-

order beliefs, dislike: 0.79, 95% CI [0.60, 0.99]; egoistic SOB: 0.22, 95% CI [0.11, 0.34]. Individuals' negative perceptions of the ambassador have a greater influence on their behavioral intentions than their beliefs about the extent to which the ambassador thinks solar panels save money.

Understanding dislike of the ambassador

To better understand the role that disliking the ambassador played in influencing people's behavioral intentions, we separately analyzed the three component questions: 1) How hypocritical is Jonathan? 2) How annoyed would you be after talking to Jonathan? 3) How much do you admire Jonathan? When combined with the other two questions, the question about admiring the ambassador was reverse-coded so that higher values would consistently indicate disliking the ambassador.

We ran a multiple regression including the consistency of the ambassador's behavior, the second-order belief components, and the ambassador attitude components, splitting up dislike into "admire", "annoyed", and "hypocrisy" (Table 2.4). All three dislike variables had a significant effect on people's intentions to adopt the behavior (*admire*: $\beta = 0.39$, $t(583) = 8.50$, $p < .001$; *annoyed*: $\beta = -0.11$, $t(583) = -2.70$, $p = 0.007$; *hypocrisy*: $\beta = -0.09$, $t(583) = -2.05$, $p = 0.04$). The extent to which people admire the ambassador had the largest effect, followed by how annoyed people were and how hypocritical they thought the ambassador was.

Consistency and inconsistency relative to the control

We also looked at the impact of the ambassador's behavior on the participants' behavioral intentions as a function of behavioral consistency, aiming to replicate Kraft-Todd et al. (2018)'s results. Participants who imagined interacting with an ambassador who behaved consistently reported a higher likelihood of purchasing solar panels through Solarize relative

Table 2.4: All three variables making up the “dislike” component (admire, annoyed, hypocrisy) had a significant negative impact on people’s behavioral intentions

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	5.34***	0.55	4.26	–	6.41	<0.001
Consistent	0.03	0.17	–0.30	–	0.37	0.844
Social Norms SOB	0.02	0.06	–0.09	–	0.14	0.691
Egoistic SOB	0.14*	0.07	0.01	–	0.28	0.036
Altruistic SOB	0.04	0.08	–0.11	–	0.19	0.583
Admire (Reverse-coded)	–0.39***	0.05	–0.48	–	–0.30	<0.001
Annoyed	–0.11**	0.04	–0.20	–	–0.03	0.007
Hypocrisy	–0.09*	0.04	–0.18	–	–0.00	0.041
Change Values	0.00	0.07	–0.14	–	0.14	0.974
Judge	0.01	0.04	–0.07	–	0.10	0.766
Judge Gap	0.01	0.08	–0.13	–	0.16	0.848
Observations	594					
R^2 / R^2 adjusted	0.387 / 0.376					
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$						

to participants who found out that the ambassador did not have solar panels on his own home, $\beta = 1.30$, $t(592) = 9.33$, $p < .001$. We then compared the effect of the presence or absence of solar panels to the control condition, in which participants were not told whether the ambassador had solar panels or not. The magnitude of the negative effect of behavioral inconsistency was larger than the magnitude of the positive effect of behavioral consistency on people’s intentions to install solar panels through the Solarize program, consistent: $\beta = 0.47$, $t(391) = 2.48$, $p = 0.01$; inconsistent: $\beta = -0.83$, $t(397) = -4.11$, $p < .001$.

We considered the possibility that the process influencing people’s behavioral intentions differed depending on the consistency or inconsistency of the ambassador. We ran separate regressions comparing the control condition to either the consistent or inconsistent conditions. When the ambassador behaved consistently, the extent to which participants admired him was the primary driver of their behavioral intentions, followed by how annoyed they were (*admire*: $\beta = 0.39$, $t(382) = 7.91$, $p < .001$; *annoyed*: $\beta = -0.15$, $t(382) = -2.54$, $p = 0.01$; *hypocrisy*:

$\beta = -0.02$, $t(382) = -0.39$, $p = 0.70$). When the ambassador behaved inconsistently, the same was true (*admire*: $\beta = 0.33$, $t(388) = 5.61$, $p < .001$; *annoyed*: $\beta = -0.12$, $t(388) = -2.30$, $p = 0.02$). However, in this case, hypocrisy also had a significant negative effect on behavioral intentions (*hypocrisy*: $\beta = -0.13$, $t(388) = -2.37$, $p = 0.02$).

In terms of second-order beliefs, none of the second-order belief components were significant when comparing behavioral intentions in the control condition to the consistent condition (*social norms*: $\beta = -0.05$, $t(382) = -0.74$, $p = 0.46$; *egoistic*: $\beta = 0.14$, $t(382) = 1.61$, $p = 0.11$; *altruistic*: $\beta = 0.06$, $t(382) = 0.59$, $p = 0.55$). However, when comparing the inconsistent condition to the control, egoistic second-order beliefs did play a role (*social norms*: $\beta = 0.08$, $t(388) = 0.96$, $p = 0.34$; *egoistic*: $\beta = 0.20$, $t(388) = 2.35$, $p = 0.02$) but not altruistic second-order beliefs (*altruistic*: $\beta = -0.03$, $t(388) = -0.32$, $p = 0.75$).

Analyzing appeal type as a robustness check

We looked at the interaction between the behavioral consistency of the advocate and whether an egoistic or altruistic reason was given. For the sake of simplicity, we did not include the combined egoistic-altruistic condition in these analyses. The main effect of consistency was still significant, $\beta = 1.42$, $t(395) = 6.00$, $p < .001$. However, there was not a significant interaction between behavioral consistency and message type, $\beta = -0.49$, $t(395) = -1.47$, $p = 0.14$.

We ran separate regressions including the ambassador's behavior, the second-order beliefs components, and the ambassador attitude components, splitting the data based on the appeal-type used: environment, money, or both (Table 2.5). This was intended to give a clearer understanding of whether the reason given by the ambassador affects the process by which the ambassador's consistency affects people's behavioral intentions. Regardless of the appeal used, disliking the ambassador was the largest driver of people's behavioral intentions, *egoistic*: $\beta = -0.66$, $t(189) = -6.23$, $p < .001$; *altruistic*: $\beta = -0.49$, $t(192) = -4.97$,

$p < .001$; *combined*: $\beta = -0.50$, $t(186) = -4.53$, $p < .001$. When splitting the data in this way, none of the second-order beliefs components were significant. One of the few differences is that people's second-order beliefs regarding social norms have a weak effect when the altruistic reason is given, *social norms SOB*: $\beta = 0.20$, $t(192) = 1.82$, $p = 0.07$. This suggests that people's perception of the ambassador's beliefs about the community norms around solar panels matters more when environmental concerns are emphasized.

Solarize vs. solar panels in general

We found that people were significantly more likely to install solar panels through Solarize, the company the ambassador represents, after interacting with an ambassador who behaved consistently relative to one who did not ($\beta = 0.65$, $t(592) = 9.33$, $p < .001$). If this is because people directly react to the hypocritical behavior of the ambassador, the reluctance to purchase should only extend to purchase options affiliated with the ambassador. If this is due to learning from inferred second order beliefs, people could either infer that there is a flaw with the panels at Solarize, or they could possibly generalize beyond Solarize and question the value of solar power at other companies as well. We checked to see whether these findings would extend to people's intentions to install solar panels in general, asking people "How likely would you be to purchase solar panels through a program other than Solarize?"

We found that people who saw an inconsistent ambassador were more likely to install solar panels through a company other than Solarize than those who saw a consistent ambassador ($\beta = 0.31$, $t(592) = 2.55$, $p = 0.01$). This suggests that the ambassador's consistency primarily influences people's perceptions of Solarize and does not seem to have a negative impact on people's willingness to install solar panels in general. When comparing people who saw an inconsistent ambassador to the control group where the ambassador's behavior was not specified, the effect of inconsistency on general behavioral intentions was not significant ($\beta = 0.19$, $t(297) = 1.08$, $p = 0.28$). Seeing a Solarize ambassador without solar panels did

Table 2.5: Disliking the ambassador is the greatest predictor of intentions to install solar panels for all three appeal types used

(a) Altruistic appeal

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	5.28***	0.98	3.35	–	7.22	<0.001
Consistent	–0.06	0.29	–0.64	–	0.51	0.826
Social Norms SOB	0.20	0.11	–0.02	–	0.41	0.071
Egoistic SOB	0.02	0.12	–0.22	–	0.26	0.876
Altruistic SOB	0.12	0.14	–0.16	–	0.40	0.399
Dislike	–0.49***	0.10	–0.69	–	–0.30	<0.001
Change Values	–0.12	0.13	–0.37	–	0.12	0.330
Judge	0.03	0.07	–0.11	–	0.17	0.677
Judge Gap	–0.14	0.12	–0.37	–	0.09	0.241
Observations	201					
<i>R</i> ² / <i>R</i> ² adjusted	0.375 / 0.349					
* <i>p</i> < 0.05 ** <i>p</i> < 0.01 *** <i>p</i> < 0.001						

(b) Egoistic appeal

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		p	
Intercept	4.71 ^{***}	0.91	2.92	–	6.50	<0.001
Consistent	–0.66 [*]	0.28	–1.22	–	–0.10	0.021
Social Norms SOB	–0.07	0.11	–0.29	–	0.15	0.546
Egoistic SOB	0.18	0.12	–0.06	–	0.42	0.148
Altruistic SOB	0.09	0.14	–0.19	–	0.37	0.536
Dislike	–0.66 ^{***}	0.11	–0.87	–	–0.45	<0.001
Change Values	0.03	0.12	–0.20	–	0.26	0.791
Judge	0.18 [*]	0.08	0.03	–	0.33	0.021
Judge Gap	0.09	0.13	–0.17	–	0.35	0.514
Observations	198					
R^2 / R^2 adjusted	0.325 / 0.296					
[*] p < 0.05 ^{**} p < 0.01 ^{***} p < 0.001						

Table 2.5: Disliking the ambassador is the greatest predictor of intentions to install solar panels for all three appeal types used (continued)

(c) Combined egoistic and altruistic appeal

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	4.74***	0.96	2.85	–	6.64	<0.001
Consistent	0.18	0.28	–0.38	–	0.74	0.530
Social Norms SOB	0.06	0.10	–0.14	–	0.25	0.552
Egoistic SOB	0.20	0.12	–0.04	–	0.44	0.108
Altruistic SOB	0.04	0.13	–0.21	–	0.29	0.742
Dislike	–0.50***	0.11	–0.71	–	–0.28	<0.001
Change Values	0.17	0.14	–0.10	–	0.45	0.218
Judge	–0.01	0.08	–0.17	–	0.14	0.872
Judge Gap	–0.20	0.15	–0.50	–	0.10	0.200
Observations	195					
<i>R</i> ² / <i>R</i> ² adjusted	0.430 / 0.406					
*<i>p</i> < 0.05 **<i>p</i> < 0.01 ***<i>p</i> < 0.001						

not taint people’s perceptions of solar panels at other companies.

We then compared the psychological mechanisms underlying people’s behavioral intentions to install solar panels through Solarize relative to their intentions to install solar panels in general. We regressed people’s intentions to purchase solar panels in general on the consistency of the advocate’s behavior, their second-order beliefs about the ambassador, their attitudes toward the ambassador, and their perceptions of how the ambassador was relating to them. While there was a significant negative effect of consistency on people’s behavioral intentions, none of the mediators were significant.

2.3.3 Discussion

We find that the extent to which people dislike the ambassador is more predictive of their behavioral intentions than their views about his beliefs or their perceptions of the ambassador judging or aiming to persuade them. Regardless of whether the behavior is framed egoistically,

altruistically, or using both types of appeals, disliking the ambassador was the strongest predictor of people’s intentions to install solar panels through Solarize. The effect of disliking the ambassador is primarily driven by people’s lack of admiration for the inconsistently-behaving ambassador as opposed to the other variables included in the “dislike” component, specifically how annoyed participants were and how hypocritical they found the ambassador.

While Kraft-Todd et al. (2018) focused on second-order beliefs as the primary and only mediator of the effect of the ambassador’s consistency on people’s behavioral intentions, we see evidence for a more nuanced interpretation. Even when second-order beliefs do influence behavior, this is primarily driven by people’s views on the ambassador’s beliefs about the money-saving benefits of the solar panels. People’s views on the ambassador’s beliefs about how much solar panels will help the environment or align with the social norms of the community did not significantly influence their behavioral intentions.

2.4 Study 2

2.4.1 *Methods*

We aim to explore the psychological mechanisms underlying people’s behavioral intentions in response to a consistent or inconsistent ambassador in a new domain; vegetarianism. Similarly to installing solar panels, there are both self-interested and prosocial motives for engaging in vegetarianism. However, this behavior also differs in important ways. It is easier financially, in that it costs less than installing solar panels. However, it might be more challenging in terms of time and effort, as it requires people to form new habits.

In this study, participants read a hypothetical message from an ambassador of a program encouraging people to go vegetarian for a month, the “Veg-for-a-Month” program through the organization GoVeg. They were asked to imagine the ambassador behaving consistently or inconsistently with the pro-vegetarian message. The ambassador’s observed behavior was

manipulated across conditions by asking people to imagine walking by the ambassador while he was eating outside at a restaurant with his family. They were told that he was eating either a veggie burger (consistent) or a bacon cheeseburger (inconsistent).

As in the previous study, we also manipulated the appeal type the ambassador used in the promotional message. After reading a description of the “Veg-for-a-Month” program ambassador, participants saw a scenario in which the ambassador used one of the following appeals.

Egoistic appeal: “Going vegetarian can help you improve your health. A vegetarian diet can improve heart health and reduce the risk of obesity and various types of cancer. Choosing to stop eating meat can make you healthier and extend your life.”

Altruistic appeal: “Going vegetarian can help you improve animals’ lives. In factory farms, animals are subject to overcrowding, exposure to disease, and high levels of stress up until they are slaughtered. Choosing to stop eating meat is a way to stop supporting this industry.”

Combined egoistic and altruistic appeal (with order of reasons counterbalanced): “Going vegetarian can help you improve your health and animals’ lives. A vegetarian diet can improve heart health and reduce the risk of obesity and various types of cancer. In factory farms, animals are subject to overcrowding, exposure to disease, and high levels of stress up until they are slaughtered. Choosing to stop eating meat is a way to stop supporting this industry, and it can make you healthier and extend your life.”

Overall, this study used a 2×3 between subjects design, crossing the ambassador’s behavior (consistent or inconsistent) with the appeal type (egoistic, altruistic, or a combination of egoistic and altruistic).

Participants were then asked to evaluate, using 1-7 Likert scales, how likely they would be to go vegetarian for a month and how likely they would be to sign up for the specific “Veg-for-a-Month” program that the ambassador is promoting. In addition to the measures of their behavioral intentions, they were asked about their beliefs about the ambassador’s beliefs (i.e., their second order beliefs), their personal attitudes toward the ambassador, and their perceptions of how the ambassador would perceive them.

Participants were recruited using Amazon Mechanical Turk. This study was pre-registered (https://aspredicted.org/JKD_731). Of the 602 participants who completed the survey, 598 passed the attention check and had unique IP addresses.

2.4.2 Results

Dimensionality reduction

We adapted the second-order beliefs scale (Table 2.6) and questions about people’s perceptions of the ambassador to relate to vegetarianism (Table 2.7). Using these updated questions, we ran principal component analyses to identify the main components. We used these to create components that map closely to the ones used in Study 1.

Assessing mediators

We regressed peoples’ intentions to try vegetarianism for a month on the ambassador’s manipulated behavioral consistency, the aggregated second-order beliefs, the components related to attitudes toward the ambassador, and the components about people’s perceptions of how the ambassador was judging or attempting to persuade them. The aggregated second-order beliefs were not significant, suggesting that they do not serve as a mediator of the ambassador’s behavior on people’s behavioral intentions in this study, $\beta = 0.08$, $t(443) = 0.95$, $p = 0.34$. The extent to which people dislike the ambassador was the largest predictor of

Table 2.6: Vegetarianism second-order beliefs

Component	Questions making up this component
	All questions begin: “How much does Jonathan believe that going vegetarian for a month. . .”
Social norms	. . . makes someone a good person in the eyes of people in your community? . . . is what people in your community think everyone should be doing? . . . helps a person’s reputation in the eyes of people in your community?
Improve health	. . . is going to help you improve your health? . . . is going to help you live longer?
Helping animals	. . . is going to help save animals’ lives? . . . is going to prevent animals from being harmed? . . . will do some part to contribute to the greater good?

Table 2.7: Vegetarian ambassador attitudes

Theoretical constructs	Component	Question
Moral hypocrisy; Amoral dislike; Admiration	Dislike	How hypocritical is Jonathan?
		How annoyed would you be after talking to Jonathan?
		How much do you admire Jonathan? (reverse-coded)
Persuasion awareness	Change your values	How likely is it that Jonathan actively wants to change your values around vegetarianism? How likely is it that Jonathan respects your values around vegetarianism? (reverse-coded)
Non-specific judgment from ambassador	Judge	If you do not go vegetarian for a month, how likely is it that Jonathan will judge you negatively?
Moral judgment from ambassador	Judgement gap	If Jonathan found out you did not go vegetarian for a month, he would think you were: 1-extremely immoral, 7-extremely moral (reverse-coded)

behavioral intentions, $\beta = -0.64$, $t(443) = -7.30$, $p < .001$.

After running a similar regression where the second-order beliefs were split up into their three components, all three measures of second-order beliefs approached significance, *egoistic SOB*: $\beta = 0.12$, $t(441) = 1.81$, $p = 0.07$, *social norms SOB*: $\beta = 0.13$, $t(441) = 1.85$, $p = 0.07$; *altruistic SOB*: $\beta = -0.15$, $t(441) = -1.94$, $p = 0.05$. People's expectations about the ambassador's beliefs about how much vegetarianism aligned with social norms, improved health, and helped animals all played a role in their behavioral intentions. However, disliking the ambassador had a greater effect than any of the second-order beliefs components, $\beta = -0.67$, $t(441) = -7.39$, $p < .001$. As in Study 1, the magnitude of the negative effect of disliking the ambassador outweighed the magnitude of the effects of any of the components of second-order beliefs (Table 2.8).

Table 2.8: Disliking the ambassador predicted people's behavioral intentions more so than second-order beliefs

Predictors	Intentions to go vegetarian through GoVeg					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	7.79***	0.79	6.25	–	9.33	<0.001
Consistent	−0.97**	0.35	−1.66	–	−0.28	0.006
Social Norms SOB	0.13	0.07	−0.01	–	0.26	0.066
Egoistic SOB	0.12	0.07	−0.01	–	0.25	0.071
Altruistic SOB	−0.15	0.08	−0.30	–	0.00	0.053
Dislike	−0.67***	0.09	−0.85	–	−0.49	<0.001
Change Values	−0.28**	0.10	−0.49	–	−0.08	0.006
Judge	0.02	0.07	−0.11	–	0.15	0.795
Judge Gap	−0.09	0.12	−0.32	–	0.14	0.437
Observations	450					
<i>R</i> ² / <i>R</i> ² adjusted	0.249 / 0.235					
* <i>p</i> < 0.05 ** <i>p</i> < 0.01 *** <i>p</i> < 0.001						

Mediation model

When running a parallel mediation model with egoistic second-order beliefs and the extent to which people dislike the ambassador, we find evidence for a complementary mediation, *total effect*: 1.19, 95% CI [0.81, 1.58], $p < 0.001$; *direct effect*: -1.29, 95% CI [-1.94, -0.64], $p < 0.001$ (Zhao et al., 2010). While the total effect is positive before the mediators are included, with the ambassador’s consistency positively impacting people’s behavioral intentions, it flips to negative after the mediators are added. Like in Study 1, disliking the ambassador mediates more of the effect than people’s beliefs about whether the ambassador believes that vegetarianism will help them improve their health, *dislike*: 2.33, 95% CI [1.74, 2.97], *egoistic SOB*: 0.14, 95% CI [-0.07, 0.37].

Understanding disliking the ambassador

We split the component that indicated disliking the ambassador into how much people admire the ambassador, how annoyed they were by him, and how hypocritical they found him. We regressed people’s behavioral intentions on these variables in addition to the ambassador’s behavioral consistency, second-order belief components, and the ambassador attitude components other than “dislike” (Table 2.9). As in the previous study, of the three variables that made up disliking the ambassador, the reduction in admiring the ambassador when behaving inconsistently had the largest effect, followed by how annoyed they were (*admire*: $\beta = 0.53$, $t(439) = 6.93$, $p < 0.001$; *annoyed*: $\beta = -0.21$, $t(439) = -3.00$, $p = 0.003$). People’s perceptions of the advocate’s hypocrisy did not predict their behavioral intentions in this study (*hypocrisy*: $\beta = 0.09$, $t(439) = 1.03$, $p = 0.30$).

Analyzing appeal type as a robustness check

We reran the regressions that included the ambassador’s consistency, second-order belief components, and ambassador attitude components, splitting the data by the appeal type used

Table 2.9: Admiration and annoyance toward the ambassador had the greatest impact on behavioral intentions among variables making up the “dislike” component

Predictors	Intentions to go vegetarian through GoVeg					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	7.70 ^{***}	0.82	6.09	– 9.30	<0.001	
Consistent	–0.32	0.40	–1.10	– 0.46	0.416	
Social Norms SOB	0.08	0.07	–0.05	– 0.21	0.228	
Egoistic SOB	0.12	0.06	–0.01	– 0.25	0.068	
Altruistic SOB	–0.15 [*]	0.07	–0.30	– 0.00	0.044	
Admire (R)	–0.53 ^{***}	0.08	–0.68	– –0.38	<0.001	
Annoyed	–0.21 ^{**}	0.07	–0.34	– –0.07	0.003	
Hypocrisy	0.09	0.09	–0.08	– 0.27	0.302	
Change Values	–0.27 ^{**}	0.10	–0.47	– –0.07	0.008	
Judge	–0.02	0.07	–0.15	– 0.11	0.722	
Judge Gap	–0.06	0.11	–0.28	– 0.16	0.601	
Observations	450					
<i>R</i> ² / <i>R</i> ² adjusted	0.289 / 0.273					
			[*] <i>p</i> < 0.05	^{**} <i>p</i> < 0.01	^{***} <i>p</i> < 0.001	

by the ambassador. Regardless of whether the ambassador used an egoistic appeal, an altruistic appeal, or a combination of both appeal types, disliking the ambassador was still the strongest predictor of people’s behavioral intentions, *egoistic*: $\beta = -0.53$, $t(133) = -3.33$, $p = 0.001$, *altruistic*: $\beta = -1.00$, $t(118) = -5.71$, $p < .001$, *combined*: $\beta = -0.59$, $t(172) = -4.06$, $p < .001$. People’s second-order beliefs about social norms have a positive effect on behavioral intentions when the altruistic-only reason is given, $\beta = 0.28$, $t(118) = 2.24$, $p = 0.03$. We looked to see whether the absence of an interaction between the ambassador’s consistency and the appeal type found in Study 1 was unique to solar panels or if it extended to vegetarianism as well. There was a directional but non-significant interaction between the reason given and the ambassador’s behavior on participants’ behavioral intentions, $\beta = 0.89$, $t(265) = 1.82$, $p = 0.07$.

GoVeg vs. vegetarianism in general

As in Study 1, we found that people were significantly more likely to participate in the Veg-for-a-Month program after interacting with a GoVeg ambassador who behaved consistently relative to one who was seen eating a cheeseburger ($\beta = 0.60$, $t(448) = 6.09$, $p < .001$). When asking people “How likely would you be to go vegetarian for a month?”, we found similar results. People were more likely to go vegetarian after interacting with a consistent advocate relative to an inconsistent one ($\beta = 0.55$, $t(448) = 5.85$, $p < .001$), even if they were not planning to engage with the GoVeg program. We also investigated the psychological mechanism at play, regressing people’s intentions to go vegetarian on the second-order belief and ambassador attitude components. The more people disliked the ambassador, the less likely they were to try vegetarianism ($\beta = -0.55$, $t(441) = -6.22$, $p < .001$). People were also less likely to engage in vegetarianism when they thought the ambassador was trying to change their values ($\beta = -0.29$, $t(441) = -2.92$, $p = 0.004$). People were more likely to engage in vegetarianism when they saw the ambassador as viewing vegetarianism as socially normative or as beneficial for health (*social norm second-order beliefs*: $\beta = 0.13$, $t(441) = 1.99$, $p < 0.05$; *egoistic second-order beliefs*: $\beta = 0.13$, $t(441) = 1.98$, $p < 0.05$).

In the case of vegetarianism, it seems that people did not strongly distinguish between the GoVeg program and trying vegetarianism more broadly, both in terms of their behavioral intentions and the psychological mechanism at play. This differed from the previous study, in which people’s intentions to install through Solarize differed from their intentions to install solar panels. This could be due to differences between the two behaviors. Installing solar panels through Solarize is a large investment, so the company from which one orders could make a big difference financially. In contrast, going vegetarian is less likely to be impacted by whether one engages with the GoVeg program or not, as people will be purchasing and preparing their own foods in either case.

2.4.3 Discussion

The effect of the ambassador’s consistency or inconsistency on people’s behavioral intentions extends to vegetarianism. As in Study 1, people are more likely to adopt a behavior after seeing that the person advocating for it engages in that behavior himself. In terms of the mechanism at play, the effect of people’s beliefs about the ambassador’s views on vegetarianism matters less than how much they like the ambassador. While people’s views on whether the ambassador genuinely believes vegetarianism is good for health do impact their behavioral intentions, this effect is relatively weak. By looking at peoples’ behavioral intentions in the context of vegetarianism, we extend the results of Study 1 to a new domain. This shows that the prior results are not unique to solar panels. More broadly, the mediating effect of disliking the ambassador stands in the case of a repeated behavior, not just a one-shot behavior like purchasing solar panels.

2.5 Study 3

2.5.1 Methods

We investigate the extent to which people’s assumptions about the ambassador shift when they are exposed to other indicators of his true beliefs beyond the observed action. This allows us to better evaluate the robustness of the prior results and see how dependent they are on the ambassador’s behavior being the only available cue of his views. Like in Study 1, participants were presented with a scenario about a hypothetical ambassador promoting solar panel purchases through the Solarize program. The main way in which this study differed from the previous studies is through an addition to the original scenario, designed as a manipulation of the ambassador’s true beliefs about solar panels. Across all conditions, participants were given a verbal indication of the ambassador’s true beliefs about solar panels, showing that he supports them when in private. After imagining their conversation with

the ambassador, which follows the same format as in Study 1, they were told the following, adapted from Kraft-Todd et al. (2018):

“Imagine that later that day, you happen to find yourself behind Jonathan in the checkout line at the grocery store. You overhear him talking on the phone, and though you do not intend to eavesdrop, your close proximity makes this inevitable. You hear him say to his wife: ‘I’m trying to do a good job as solar ambassador on this campaign. Solar panel technology has improved a lot recently.’”

As in study 1, the ambassador’s behavioral consistency was manipulated; the ambassador was described as either having solar panels, not having solar panels, or it was not specified. For robustness, the appeal type the ambassador used was either egoistic, altruistic, or a combination of both egoistic and altruistic reasons. Depending on the appeal type being used, the ambassador mentioned one of the three following reasons for installing solar panels during the overheard conversation with his wife, adapted from Kraft-Todd et al. (2018).

Egoistic: “I think the amount of time it will take to pay themselves off has never been shorter!”

Altruistic: “I think the panels have a real potential to make the planet greener!”

Combined egoistic and altruistic: “I think the amount of time it will take to pay themselves off has never been shorter, and I think the panels have a real potential to make the planet greener!”

Overall, this study used a 3×3 between-subjects design, which crossed the ambassador’s solar panel ownership with the appeal type.

Participants were then asked to evaluate how likely they would be to install solar panels through Solarize using a 7-point Likert scale. They were also asked to evaluate their perceptions of the ambassador’s beliefs, their personal attitudes toward the ambassador, and their perceptions of the ambassador’s attitudes toward them.

In this pre-registered study (https://aspredicted.org/QJZ_PJ2), we recruited participants through Amazon Mechanical Turk. Of the 905 participants who completed the survey, 896 passed the attention check and had unique IP addresses.

2.5.2 Results

Regressions

We regressed people’s behavioral intentions on the advocate’s behavior, a measure aggregating all questions about second-order beliefs, and the four components related to people’s perceptions of the ambassador. We used the same components identified in Study 1. Both the aggregated second-order beliefs scale and disliking the ambassador predicted people’s behavioral intentions, *aggregated SOB*: $\beta = 0.37$, $t(389) = 4.65$, $p < .001$, *dislike*: $\beta = -0.43$, $t(389) = -6.51$, $p < .001$. The magnitude of disliking the ambassador was larger than the effect of inferred ambassador beliefs about solar panels.

When splitting up the second-order beliefs scale into the three components based on the prior PCA results (Table 2.10), only egoistic second order beliefs significantly predicted people’s intentions to install solar panels, $\beta = 0.14$, $t(387) = 2.30$, $p = 0.02$. In addition to people’s views on the advocate’s beliefs about the money-saving benefits of solar panels, the extent to which they disliked him also impacted their behavioral intentions, $\beta = -0.43$, $t(387) = -6.32$, $p < .001$. Their attitudes toward the ambassador mattered more than their views on his beliefs about the selfish benefits of installing solar panels.

Mediation model

Together, people’s egoistic second-order beliefs and their dislike of the ambassador fully mediate the effect of the ambassador’s behavior on their own behavioral intentions, total effect: 0.34, 95% CI [0.06, 0.61], $p = 0.01$; direct effect: -0.08, 95% CI [-0.33, 0.17], $p = 0.54$.

Table 2.10: Disliking the ambassador predicted people’s behavioral intentions more so than second-order beliefs. Among second-order beliefs, egoistic second-order beliefs had the greatest impact.

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	5.00***	0.71	3.61	–	6.39	<0.001
Consistent	–0.29	0.18	–0.64	–	0.06	0.100
Social Norms SOB	0.12	0.07	–0.03	–	0.26	0.110
Egoistic SOB	0.14*	0.06	0.02	–	0.26	0.022
Altruistic SOB	0.11	0.07	–0.04	–	0.26	0.148
Dislike	–0.43***	0.07	–0.57	–	–0.30	<0.001
Change Values	–0.24**	0.09	–0.43	–	–0.06	0.009
Judge	0.06	0.06	–0.05	–	0.17	0.286
Judge Gap	0.07	0.11	–0.14	–	0.28	0.522
Observations	396					
R^2 / R^2 adjusted	0.291 / 0.276					
*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$						

When comparing the indirect effects of each of these mediators, disliking the ambassador explains more of the mediation process than people’s views about the extent to which the ambassador thinks solar panels save money, dislike: 0.38, 95% CI [0.26, 0.51]; egoistic SOB: 0.04, 95% CI [0.00, 0.09].

Understanding disliking the ambassador

We split the component about disliking the ambassador into the three variables that made it up; admiration for the ambassador, annoyance at the ambassador, and perceptions of the ambassador’s hypocrisy (Table 2.11). We ran a similar regression to the ones above, regressing people’s behavioral intentions on the ambassador’s behavioral consistency, second-order belief components, ambassador attitude components, and the three separate variables that made up people’s dislike of the ambassador. Whether they admired the ambassador was the largest predictor of people’s behavioral intentions. Among the three variables making up dislike,

this was followed by annoyance and hypocrisy (*admire*: $\beta = 0.39$, $t(385) = 6.18$, $p < 0.001$; *annoyed*: $\beta = -0.24$, $t(385) = -4.31$, $p < 0.001$; *hypocrisy*: $\beta = 0.11$, $t(385) = 2.18$, $p = 0.03$).

Table 2.11: Among the components making up “dislike”, admiration and annoyance were the largest predictors of people’s behavioral intentions

Predictors	Intentions to install solar panels through Solarize					
	Estimates	std. Error	CI		<i>p</i>	
Intercept	5.86 ^{***}	0.74	4.40	–	7.31	<0.001
Consistent	0.05	0.18	–0.31	–	0.41	0.796
Social Norms SOB	0.13	0.07	–0.01	–	0.27	0.079
Egoistic SOB	0.08	0.06	–0.04	–	0.19	0.179
Altruistic SOB	0.06	0.07	–0.09	–	0.20	0.439
Admire (R)	–0.39 ^{***}	0.06	–0.51	–	–0.26	<0.001
Annoyed	–0.24 ^{***}	0.06	–0.35	–	–0.13	<0.001
Hypocrisy	0.12 [*]	0.05	0.01	–	0.22	0.030
Change Values	–0.24 ^{**}	0.09	–0.42	–	–0.06	0.008
Judge	0.04	0.05	–0.07	–	0.15	0.487
Judge Gap	0.06	0.10	–0.15	–	0.26	0.585
Observations	396					
<i>R</i> ² / <i>R</i> ² adjusted	0.348 / 0.331					
		[*] <i>p</i> < 0.05	^{**} <i>p</i> < 0.01	^{***} <i>p</i> < 0.001		

We split up the data in order to run similar regressions comparing the control to the consistent condition and the control to the inconsistent condition. The results suggest that the process driving peoples’ behavioral intentions is similar when the ambassador is consistent vs. when he is inconsistent. In both cases, the extent to which people admire the ambassador and the extent to which they find him annoying are the primary drivers of people’s behavioral intentions (consistent: *admire*: $\beta = 0.41$, $t(359) = 6.36$, $p < 0.001$; *annoyed*: $\beta = -0.20$, $t(359) = -3.26$, $p = 0.001$; inconsistent: *admire*: $\beta = 0.25$, $t(371) = 3.70$, $p < 0.001$; *annoyed*: $\beta = -0.22$, $t(371) = -3.82$, $p < 0.001$). In the case where the ambassador behaves consistently, admiration matters more relative to how annoyed people are than in the case where the ambassador behaves inconsistently.

Consistency and inconsistency relative to the control

Replicating the results from the previous two studies and Kraft-Todd et al. (2018), people are more likely to report higher intentions to install solar panels when they see the solar panel ambassador has panels on his house relative to when he does not, $\beta = 0.63$, $t(394) = 3.96$, $p < .001$. We compared the effect of behavioral consistency and inconsistency to a control condition. This control condition differed from the control condition in Study 1. In this study, it is not specified whether the ambassador has panels, but participants still saw one of the three possible appeal types. In contrast, in study 1, the control condition did not include information about the ambassador's solar panel ownership, and there was no specific reason given.

In this study, there was no effect of the ambassador's behavioral consistency on participants' behavioral intentions compared to the control condition, $\beta = -0.001$, $t(368) = -0.008$, $p = 0.99$. When comparing the effect of inconsistency (the ambassador not having solar panels) to the control, however, there was a significant negative effect on people's intentions to purchase solar panels compared to the control condition, $\beta = -0.63$, $t(380) = -4.06$, $p < .001$.

When the ambassador behaved consistently, the extent to which participants admired him was the primary driver of their behavioral intentions, followed by how annoyed they were (*admire*: $\beta = 0.41$, $t(359) = 7.91$, $p < .001$; *annoyed*: $\beta = -0.20$, $t(359) = -3.26$, $p = 0.001$; *hypocrisy*: $\beta = 0.03$, $t(359) = 0.50$, $p = 0.62$). When the ambassador behaved inconsistently, the same was true (*admire*: $\beta = 0.25$, $t(371) = 3.70$, $p < .001$; *annoyed*: $\beta = -0.22$, $t(371) = -3.82$, $p < 0.001$). In this case, unlike in Study 1, hypocrisy did not have a significant negative effect on behavioral intentions (*hypocrisy*: $\beta = 0.07$, $t(371) = 1.22$, $p = 0.22$).

Appeal type

We checked to see whether the relevance of people’s dislike of the ambassador varied depending on the type of appeal the ambassador gave. We ran regressions of people’s behavioral intentions on the advocate’s behavior, second-order belief components, and attitudes toward the advocate, splitting the data based on whether the ambassador gave an egoistic appeal, altruistic appeal, or combination of the two. In all three cases, the degree to which people like or dislike the ambassador is the main variable impacting their own behavioral intentions, *egoistic*: $\beta = -0.38$, $t(114) = -3.02$, $p = 0.003$; *altruistic*: $\beta = -0.44$, $t(103) = -3.13$, $p = 0.002$; *combined*: $\beta = -0.45$, $t(152) = -4.35$, $p < .001$. When the ambassador emphasized the environmental benefits of installing solar panels (altruistic appeal type), people’s perceptions of the ambassador’s views of the norms around having panels also predicted their behavioral intentions, $\beta = 0.35$, $t(103) = 2.32$, $p = 0.02$.

2.5.3 Discussion

The impact of the ambassador’s behavior on people’s behavioral intentions follows a similar pattern as in the previous two studies, despite the manipulation which provided an additional cue of the ambassador’s true beliefs. As in the previous two studies, disliking the ambassador is the primary predictor of people’s behavioral intentions. People’s egoistic second-order beliefs also have a significant effect. When breaking down dislike into its separate components, the extent to which people admire the ambassador and are annoyed by him are the primary drivers of behavioral intentions. When splitting up the second-order beliefs about the ambassador, people’s views about the ambassador’s beliefs about the money-saving benefits of installing solar panels influence their behavior more than their views on the ambassador’s beliefs about the environmental benefits or social norms.

By adding another possible signal of the ambassador’s true beliefs through the overheard conversation, we gained a better understanding of the extent to which the previous effects

are driven by people making inferences about true beliefs through the ambassador’s actions. The fact that the impact of behavioral consistency and inconsistency replicates despite the ambassador privately stating his beliefs highlights the stronger influence of actions over words. Regardless of whether the ambassador conveys consistent personal beliefs in private, people do not like someone who behaves in seemingly contradictory ways. These results align with the comparable conditions of experiment 2 in Kraft-Todd et al. (2018); when the ambassador privately expresses positive beliefs, they also found that there is still a positive effect of ambassador behavioral consistency on people’s intentions to install solar panels.

While Study 3 (with access to a cue of true beliefs) largely replicated Study 1 (in which no such cue was provided), there was one notable difference in results. In Study 1, there was a positive effect of the advocate’s behavioral consistency on people’s intentions to install solar panels relative to the condition where his behavior was not specified. In contrast, that effect was not observed in this study. While there are multiple potential interpretations of this, one possibility is that, in the previous study, the ambassador’s behavior was the only indication of how he privately felt about the solar panels he was supporting. Therefore, it may have had a greater effect relative to having no information at all. In the current study, participants have an indication of his private beliefs in all conditions through overhearing the conversation with his wife. Having panels does not seem to add much to that signal, but not having panels makes people question it. The large negative effect of the ambassador not having panels on people’s interest in installing panels themselves shows that his words with his wife do not make up for him not having panels himself; actions that show a lack of support matter more than words that do indicate support.

2.6 General Discussion

We explored how an advocate’s behavior influences people’s motivation to adopt similar behaviors. In three studies, we gained a clearer understanding of the mechanisms behind this

process in the case of behaviors that benefit the individual performing them and society.

Among the possible mediators we measured, we found that the extent to which people disliked an advocate was the primary determinant of whether they wanted to engage in a new behavior themselves. This played a larger role than people's inferences about the ambassador's beliefs, their concerns about being judged by the ambassador, and their resistance to a persuasion attempt. This effect was so robust that it held even when splitting the data based on whether people were shown the egoistic, altruistic, or combined appeal type.

Disliking the ambassador was a better predictor of people's behavioral intentions than people's inferences about the ambassador's beliefs. Kraft-Todd et al. (2018) posited that people are more likely to install solar panels on their own home after interfacing with a solar panel advocate who owns solar panels compared to one who does not, claiming that the advocate's solar panel ownership is a believable signal of value. We used the second-order beliefs scale developed by Kraft-Todd et al. (2018) and found that disliking the ambassador outweighed the impact of second-order beliefs in all three studies.

These results align with prior literature demonstrating the ways in which role models can inspire people to change their behavior (Lockwood & Kunda, 2000; Thrash & Elliot, 2003). They also align with work showing that people are less likely to purchase products from a company when they do not like the celebrity spokesperson associated with it (Min et al., 2019). People are more likely to support an organization or engage in a behavior when they like the person promoting it. People's perceptions of the ambassador's hypocrisy did not play a significant role in determining their behavioral intentions. This allowed us to rule out theories that suggested that perceptions of hypocrisy would make people less likely to behave prosocially.

Disliking the ambassador also had a greater impact on behavioral intentions than people's concerns about how the ambassador was relating to them, in particular whether the ambas-

sador would judge them negatively or was presenting misleading information with the primary intention to persuade them. While social pressure has repeatedly proven to be a powerful motivator, it did not stand out as a mediator in our studies. There are several potential ways of interpreting these findings. One is that the scenarios used were not adequate to convey the sense of social pressure someone might feel when face to face with a member of their community who they know they will see in the future. Alternatively, social pressure may not play a significant role in the case of non-normative (not widely adopted) prosocial behaviors like solar panel installation and vegetarianism. Given that most community members will not have adopted these behaviors, the threat of judgment from the ambassador may not be too concerning. Similarly, the concern about a persuasion attempt may be less likely to elicit reactance when the ambassador is the only person exerting pressure. It may be worth continuing to explore these psychological constructs in instances where there is a stronger societal expectation for a person to adopt a behavior.

There is a common thread connecting the dominance of dislike over second-order beliefs and people's perceptions of the ambassador's attitudes toward them. The CRED theory, concerns about judgment, and persuasion awareness all assume a large degree of inferential effort. They assume that when a person notices an ambassador behaving consistently or inconsistently, their primary mental action will involve considering how an ambassador thinks, either about the action being promoted or about them. Disliking an inconsistent ambassador is likely a much quicker, easier mental process than engaging in a deliberative process about what the ambassador might be thinking. Disliking and disengaging from someone whose words and deeds do not match could be a simple mental heuristic, allowing people to protect themselves in a low-effort way. It may be worth further exploring when people are likely to engage in careful deliberations in the face of inconsistency and if simpler aversive reactions hold true in other settings.

This simple-heuristic explanation is consistent with our other findings. When including a

control group in which the solar ambassador’s behavior was not specified, we found that the magnitude of inconsistency’s effect on behavioral intentions was greater than the effect of consistency. While consistency did lead to a boost in Studies 1 and 2, it was small relative to the effect of the ambassador not owning solar panels. The effect of inconsistency is so strong that it even holds in a scenario where the ambassador is overheard having an ostensibly honest, private conversation with his wife about how much he supports solar panels. In these cases, the negative appraisal of the ambassador’s inconsistency overshadows evidence that he does value solar panels. When the ambassador does not own solar panels, it does not matter what he says about them in private; the vast majority of people are less likely to even consider solar panel installation (at least from the same company) as a result. This suggests that in cases of advocacy for behavior change, consistency of behavior is extremely important. People are unlikely to go through the effort of considering possibilities for why an advocate behaves inconsistently, instead dismissing them and preserving their mental energy for other endeavors. This could potentially extend to a wide range of other advocacy and promotion behaviors, ranging from political advocacy to sales.

2.6.1 When inferences about beliefs do matter

While dislike for an inconsistent advocate is the primary determinant of people’s behavior, we do find that certain beliefs about the ambassador’s beliefs play an important role. To better understand the role of second-order beliefs, we ran PCA analyses on Kraft-Todd et al. (2018)’s second-order beliefs scale as measured in our studies. We found that there were three distinct components at play: egoistic beliefs, altruistic beliefs, and beliefs about social norms. Egoistic beliefs represented the extent to which the advocate thought solar panels would save money or thought vegetarianism was positive for health. Altruistic beliefs represented the advocate’s views on whether the solar panels were good for the environment and whether vegetarianism was good for animals. Social norm beliefs represented the extent to which the

ambassador believed the behavior to be a social norm.

Egoistic second-order beliefs were the only component from the second-order beliefs scale that consistently predicted people’s behavioral intentions when we regressed intentions on second-order beliefs and the other aggregated components. While Kraft-Todd et al. (2018) claimed that second-order beliefs generally influence adoption of non-normative prosocial behaviors, our findings qualify these results. People’s inferences about the ambassador’s beliefs do matter, but primarily insofar as they provide information about whether the new behavior will directly benefit or harm themselves. Even when a behavior has a prosocial element, most people focus on how it will impact them. This makes sense in the case of a behavior where there is not yet a strong social norm; with the absence of a social norm in place, people are likely to think about their choice in the same way they would think about other egoistically-driven choices.

When splitting the data by the appeal type used by the ambassadors, we find a different pattern in the case of the altruistic appeal type. When the ambassador promotes solar panels for the sake of the environment or vegetarianism for the sake of the animals, people’s second-order beliefs about social norms predict their behavioral intentions. In other words, the extent to which people believe that engaging in this new behavior is a social norm impacts the extent to which people report they will engage in the behavior themselves. This suggests that when altruistic reasons are highlighted without mention of the selfish benefits, people are more likely to consider societal expectations. This shows that people are somewhat sensitive to the argument being used to persuade them, particularly when a prosocial argument is the only one used.

These results help qualify the findings from Kraft-Todd et al. (2018). While these authors group all second-order beliefs together, we find that people are sensitive to different second-order beliefs depending on the context. When looking across all appeal types used by the ambassador, people are more inclined to focus on self-interest. When someone is highlighting

that solar panels are a good financial investment or that vegetarianism is beneficial for health as at least one of the reasons why someone should engage in this behavior, the ambassador’s inconsistency casts doubt on these claims. If the ambassador is behaving inconsistently, people may assume that the ambassador has private information about whether the behaviors are actually beneficial. Alternatively, the ambassador may be highlighting the selfish benefits in order to convince people to adopt the behavior, regardless of whether these claims are true. In the absence of an egoistic appeal, however, individuals become more attuned to cues regarding societal expectations. Their perceptions of the ambassador’s view that the behavior is a social norm predict their willingness to report adopting this behavior. These findings suggest that people incorporate their inferences regarding the ambassador’s beliefs in different ways based on the framing of the appeal. Self-interest takes precedence when included; however, in the absence of self-interest, people pay more attention to the ambassador’s views of social norms.

2.6.2 Limitations and next steps

It would be worth exploring whether these results hold when looking at behaviors other than solar panel installation and vegetarianism. In particular, it would be interesting to look at the impact of dislike and second-order beliefs on people’s behavioral intentions in the case of behaviors that can be construed as only egoistic or only altruistic. While we find that an advocate’s inconsistency elicits negative reactions regardless of the appeal type used, it could be valuable to understand the ways in which these reactions differ when considering behaviors that do not blend different types of motives.

Additionally, it could be valuable to explore whether these results hold when looking at real behavior. It is possible that people’s second-order beliefs about whether the advocate saw the behavior as a social norm elicited more positive behavioral intentions not because people’s behavior would actually change, but because they felt compelled to report that their

behavior would change when a behavior was construed as being for the benefit of others. More broadly, people may have inflated their behavioral intentions, perhaps as a means of signaling their positive intentions to themselves and to researchers.

It could also be worth looking at whether there are ways to mitigate the negative effects of behavioral inconsistency in cases when an advocate or salesperson cannot engage in the behavior themselves. When promoting behaviors that are challenging or costly, someone may sincerely support those behaviors while not being able to engage in them; for example, a passionate vegetarian advocate may want to encourage a wealthy potential donor to give in ways that would be unattainable for the advocate. Similarly, an apartment-dweller with no control over their roof may want to encourage homeowners to install solar panels, despite not having the option to own panels themselves. Perhaps providing a plausible explanation for the lack of behavioral consistency would attenuate the effect of ostensible hypocrisy on people's subsequent behavior.

2.7 Conclusion

Our studies offer a more nuanced understanding of the psychological mechanisms determining people's behavioral intentions in response to word-deed alignment and misalignment. When seeing someone behave hypocritically, people are more likely to have a direct negative reaction and disengage than they are to engage in inference about the implications for that person's private beliefs. To the extent that they are thinking about an advocate's beliefs, they are primarily focused on what those beliefs mean about how they could personally be impacted. Even if an advocate has a perfectly good explanation for why they do not have solar panels or why they were eating a cheeseburger, they likely would not be given the chance to explain themselves. In a world where people are often trying to capture our attention and change our behavior, writing off people who do not practice what they preach may serve as a useful strategy of self-preservation; however, it may also promote misunderstandings, and pointing

out even trivially inconsistent behavior may be leveraged as a way to undermine the credibility of advocates.

REFERENCES

- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1), 544–555. <https://doi.org/10.1257/aer.99.1.544>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Baumeister, R. F., & Tice, D. M. (1990). Point-counterpoints: Anxiety and social exclusion. *Journal of Social and Clinical Psychology*, 9(2), 165–195. <https://doi.org/10.1521/jscp.1990.9.2.165>
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–62, Vol. 6). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60024-6](https://doi.org/10.1016/S0065-2601(08)60024-6)
- Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, 24(4), 586–607. <https://doi.org/10.1016/j.jcps.2014.05.002>
- Berkowitz, L., & Daniels, L. R. (1964). Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships. *The Journal of Abnormal and Social Psychology*, 68(3), 275–281. <https://doi.org/10.1037/h0040164>

- Bodner, R., & Prelec, D. (2003). The diagnostic value of actions in a self-signaling model. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions, vol. i: Rationality and well-being* (pp. 105–123). Oxford University Press.
- Bolderdijk, J. W., Brouwer, C., & Cornelissen, G. (2018). When do morally motivated innovators elicit inspiration instead of irritation? *Frontiers in Psychology, 8*, 2362. <https://doi.org/10.3389/fpsyg.2017.02362>
- Bourgeois, M. J., Sommer, K. L., & Bruno, S. (2009). What do we get out of influencing others? *Social Influence, 4*(2), 96–121. <https://doi.org/10.1080/15534510802465360>
- Brehm, J. W., & Sensenig, J. (1966). Social influence as a function of attempted and implied usurpation of choice. *Journal of Personality and Social Psychology, 4*(6), 703–707. <https://doi.org/10.1037/h0023992>
- Centers for Disease Control and Prevention. (n.d.). Masks and respiratory viruses prevention. Retrieved April 23, 2025, from <https://www.cdc.gov/respiratory-viruses/prevention/masks.html>
- Centers for Disease Control and Prevention. (2020, December 31). How to protect yourself & others. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*(1), 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Conjointly. (n.d.). *What is TURF analysis and when to use it? - conjointly*. Retrieved May 2, 2025, from <https://conjointly.com/blog/turf-analysis/>

- Effron, D. A., O'Connor, K., Leroy, H., & Lucas, B. J. (2018). From inconsistency to hypocrisy: When does “saying one thing but doing another” invite condemnation? *Research in Organizational Behavior*, 38, 61–75. <https://doi.org/10.1016/j.riob.2018.10.003>
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318. <https://doi.org/10.1111/j.1467-9280.2006.01704.x>
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327–339. <https://doi.org/10.1037/0022-3514.87.3.327>
- Feiler, D. C., Tost, L. P., & Grant, A. M. (2012). Mixed reasons, missedgivings: The costs of blending egoistic and altruistic reasons in donation requests. *Journal of Experimental Social Psychology*, 48(6), 1322–1328. <https://doi.org/10.1016/j.jesp.2012.05.014>
- 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>
- Feinberg, M., & Willer, R. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, 13(12), e12501. <https://doi.org/10.1111/spc3.12501>
- Fisher, R. J., Vandenbosch, M., & Antia, K. D. (2008). An empathy-helping perspective on consumers' responses to fund-raising appeals. *Journal of Consumer Research*, 35(3), 519–531. <https://doi.org/10.1086/586909>

- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, 18(2), 255–297. <https://doi.org/10.1111/0162-895x.00058>
- Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1), 1–31. <https://doi.org/10.1086/209380>
- Graham, J., Meindl, P., Koleva, S., Iyer, R., & Johnson, K. M. (2015). When values and behavior conflict: Moral pluralism and intrapersonal moral hypocrisy. *Social and Personality Psychology Compass*, 9(3), 158–170.
- Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and relevant moral exemplars are more effective than extraordinary exemplars in promoting voluntary service engagement. *Frontiers in Psychology*, 8, 283. <https://doi.org/10.3389/fpsyg.2017.00283>
- Heath, C. (1999). On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes*, 78(1), 25–62. <https://doi.org/10.1006/obhd.1999.2826>
- Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, 30(4), 244–260. <https://doi.org/10.1016/j.evolhumbehav.2009.03.005>
- Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3), 165–196. [https://doi.org/10.1016/s1090-5138\(00\)00071-4](https://doi.org/10.1016/s1090-5138(00)00071-4)
- Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin*, 69(4), 248–268. <https://doi.org/10.1037/h0025725>

- Hoogendoorn, G., Sütterlin, B., & Siegrist, M. (2019). When good intentions go bad: The biased perception of the environmental impact of a behavior due to reliance on an actor's behavioral intention. *Journal of Environmental Psychology*, 64, 65–77.
- Jordan, J., & Sommers, R. (2020, December 25). False signaling and personal moral failings: Two distinct pathways to hypocrisy with unequal moral weight. <https://doi.org/10.31234/osf.io/87txd>
- Kareklas, I., Carlson, J. R., & Muehling, D. D. (2014). “i eat organic for my benefit and yours”: Egoistic and altruistic considerations for purchasing organic food and their implications for advertising strategists. *Journal of Advertising*, 43(1), 18–32.
- Karniol, R. (2003). Egocentrism versus protocentrism: The status of self in social prediction. *Psychological Review*, 110(3), 564–580. <https://doi.org/10.1037/0033-295X.110.3.564>
- Klein, N., & O'Brien, E. (2018). People use less information than they think to make up their minds. *Proceedings of the National Academy of Sciences*, 115(52), 13222–13227. <https://doi.org/10.1073/pnas.1805327115>
- Kraft-Todd, G. T., Bollinger, B., Gillingham, K., Lamp, S., & Rand, D. G. (2018). Credibility-enhancing displays promote the provision of non-normative public goods. *Nature*, 563(7730), 245–248. <https://doi.org/10.1038/s41586-018-0647-4>
- Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, 67(4), 596–610. <https://doi.org/10.1037//0022-3514.67.4.596>
- Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, 73(1), 91–103. <https://doi.org/10.1037/0022-3514.73.1.91>

- Lockwood, P., & Kunda, Z. (2000). Outstanding role models: Do they inspire or demoralize us? In A. Tesser, R. B. Felson, & J. M. Suls (Eds.), *Psychological perspectives on self and identity* (pp. 147–171). American Psychological Association.
- Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, *74*(1), 53–62. <https://doi.org/10.1037//0022-3514.74.1.53>
- Min, J. H. J., Chang, H. J. J., Jai, T.-M. C., & Ziegler, M. (2019). The effects of celebrity-brand congruence and publicity on consumer attitudes and buying behavior. *Fashion and Textiles*, *6*(1), 1–19. <https://doi.org/10.1186/s40691-018-0159-8>
- Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*(2), 200–207. <https://doi.org/https://doi.org/10.1177/1948550611415695>
- Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 341–354). Cambridge University Press.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X)
- Shen, L., Hsee, C. K., Wu, Q., & Tsai, C. I. (2012). Overpredicting and underprofiting in pricing decisions. *Journal of Behavioral Decision Making*, *25*(5), 512–521.
- Sommer, K. L., & Bourgeois, M. J. (2010). Linking the perceived ability to influence others to subjective well-being: A need-based approach. *Social Influence*, *5*(3), 220–244.

- 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)
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870. <https://doi.org/10.1037/0022-3514.78.5.853>
- Thrash, T. M., & Elliot, A. J. (2003). Inspiration as a psychological construct. *Journal of Personality and Social Psychology*, 84(4), 871–889. <https://doi.org/10.1037/0022-3514.84.4.871>
- Weaver, K., Garcia, S. M., & Schwarz, N. (2012). The presenter's paradox [Place: US Publisher: Univ of Chicago Press]. *Journal of Consumer Research*, 39(3), 445–460. <https://doi.org/10.1086/664497>
- Weiss, A., Burgmer, P., & Mussweiler, T. (2018). Two-faced morality: Distrust promotes divergent moral standards for the self versus others. *Personality and Social Psychology Bulletin*, 44(12), 1712–1724. <https://doi.org/10.1177/0146167218775693>
- White, K., & Peloza, J. (2009). Self-benefit versus other-benefit marketing appeals: Their effectiveness in generating charitable support. *Journal of Marketing*, 73(4), 109–124. <https://doi.org/10.1509/jmkg.73.4.109>
- Wiener, J. L., & Mowen, J. C. (1986). Source credibility: On the independent effects of trust and expertise. *ACR North American Advances*, 13, 306–310.
- Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, 21(2), 101–112.

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206.
<https://doi.org/10.1086/651257>