# Conclusions

The purpose of this dissertation was to test different strategies for increasing openness to belief change, as well as to understand resistance to change and the effects of belief polarization on that resistance. To do this, I created an experiment directly testing the effectiveness of increasing or decreasing social consensus on support for a variety of polarized topics (Study 1). Study 1 demonstrated that, across a variety of topics, there was a significant interaction between the effects of a social consensus manipulation and time. Furthermore, there was evidence that a social consensus manipulation was able to significantly increase support for Universal Healthcare. I then executed on another experiment to understand the effects of moral conviction manipulation on an expanded set of polarized and non-polarized topics (Study 2). There was no evidence to support the hypotheses that the two ‘moral’ interventions would increase moral conviction and the pragmatic and hedonic interventions would decrease moral conviction, relative to the control group. While Study 1 and Study 2 examined the impact of moral conviction and social consensus individually, the purpose of Study 3 was to test the interaction between the social consensus and moral conviction manipulations on support for polarized topics. Study 3 also provided the opportunity to replicate the findings of Study 1 and Study 2. Study 3 utilized a within-subjects design that measured support for the topics before and after each of the four interventions (i.e., 2 [high or low social consensus] x 2 [moral essay or pragmatic essay]). While support for all of the topics increased post-intervention compared to pre-intervention, there was no main effect of the social consensus manipulation on support for any of the three topics (UHC, capital punishment, usage of AI in the workforce); thus the findings of Study 1 were not replicated with these topics. Additionally, there was no significant interaction between the two factors of moral conviction and social consensus on support for the topics.

# Discussion

In light of the full set of results, some preliminary conclusions seem reasonable to draw. I was not successful at empirically replicating the finding that moral conviction inoculates individuals from the effects of social consensus (Skitka, 2021). One plausible explanation is that the moral conviction manipulations were improperly designed. While previous literature has indicated that framing arguments using moral terms (e.g., freedom, liberty, etc.) or centering on perceptions of harm (e.g., harmful, dangerous, contaminated, etc.) increases perceptions of moral conviction (Kodapanakkal 2021; Clifford, 2019), I was unable to successfully replicate this pattern in Study 2 or 3. Plausibly, this is due to the fact that the topics covered in prior literature were generally seen to be unpolarized, or at least not explicitly polarized (e.g., hiring algorithms, fish farming), whereas in Study 2 and 3, the majority of the topics were explicitly polarized (UHC, capital punishment, climate change, etc.).

Additionally, I was not able to find support for the existence of an interaction between social consensus and moral conviction. One conclusion is that this potentially stems from difficulty in successfully manipulating moral conviction. Further improvements in moral conviction manipulation could plausibly allow for replication of this study with superior results. However, another explanation is due to the shared orientation of the arguments in Study 3; all of the essays were ‘in favor’ of the topic in question. A reasonable conclusion is that the directionality of the persuasive arguments (e.g., all in favor) was the most significant factor, as that interpretation falls in line with the observed results. I could investigate this further in future studies by presenting persuasive essays that are both in favor and opposition to the topic in question.

This dissertation showed mixed success at manipulating perception of social consensus using methods adapted from Kobayashi (2018). One plausible reason for the mixed success is that the topics chosen (UHC, climate change, capital punishment), diverged significantly from the original set of topics Kobayashi chose to use (climate change, blood type personality, nuclear power, and whale research). Additionally, it is important to note that while the version of the manipulation that I used for this study explicitly relied on deceiving the participants as to the base rate of consensus for each of the topics, the result that there are substantive differences in support due to shared perceptions on what is popular or unpopular should generalize broadly. Deceiving the public in order to manipulate support for a topic would generally be seen as unethical, but increasing salience of the public as to how much consensus there actually is (i.e., just because the American public supports something in general, that doesn’t mean any given individual is aware of it) could be a low-cost intervention that leverages the strengths of social consensus effects.

One direction for future work on polarized belief formation and change that could be worth exploring is to delve deeper into using domain specific information. In a pragmatic sense, one approach would be to conduct extensive qualitative research with individuals that have polarized beliefs, so as to determine which shared traits or characteristics of the topic are seen with a moral lens. Domain specific pragmatic or moral arguments would plausibly be more effective at changing moral convictions, as compared to using the ‘general’ framework of arguments based on morality and harm (increasing moral conviction) or arguments based on economics and practical implementation (decreasing moral conviction). Another context in which domain specific information could provide deeper insight is individual characteristic differences in topics themselves (i.e., why did social consensus manipulation work for UHC but not the other topics?). Future research could focus on what shared characteristics exist in topics that are or are not vulnerable to the effects of social consensus. Knowledge of those shared characteristics would help calibrate how to spend limited public outreach resources such that topics are prioritized based on how likely a social consensus intervention will be effective.

Another direction for future study would be to explore different methods for obtaining evidence of actual revealed preferences, rather than relying on self-reported support for a topic score. One trivially simple example could apply to the non-polarized topic of ‘exercise’ used in Study 2; instead of explicitly measuring ‘support for exercise’ on some form of survey, revealed preferences in the form of biometric information (e.g., step counter, activity tracker, etc.) could directly show evidence of a social consensus or moral conviction manipulation being effective at changing behavior. Another example with the more ‘polarized’ topic of climate change used in studies 2 and 3, would be revealed preferences as they relate to sustainable behavior. For example, it is relatively simple to track (by weight or volume) total amount of landfill waste and/or recycling generated by a single individual over a fixed period of time. These more direct measuresThese direct measures, however, are not well suited to tracking revealed preferences for societal level behaviors (i.e., rate of executions for capital punishment, % of population covered by universal health care).

Finally, one future goal for this research would be to see if the effect of social consensus on belief formation and change functions differently based on the type of social consensus manipulation. The manipulation I used in my research earlier was effective, but relatively impersonal. It would be very useful to research whether the effects of social consensus are greater in small or large group settings, where the relevant comparison group is not the nebulous concept of ‘Americans’ as a whole, but instead the social group immediately and physically around you. This seems especially pertinent, as this structure mimics actual human social dynamics (i.e., social consensus is assessed and formed through shared, in person experience), and thus has greater external validity.

In summary, I found mixed evidence that manipulated social consensus affects support for a variety of polarized topics; In the case where it was successful (UHC), greater perception of social consensus in favor of a topic was associated with increased support. However, there was no evidence that the obverse was associated with decreased support. Furthermore, I found that I was unable to successfully manipulate moral conviction for several polarized topics by framing arguments using moral terms (e.g., freedom, liberty, etc.) or centering on perceptions of harm (e.g., harmful, dangerous, contaminated, etc.). Given this, I could not find evidence in support of a significant interaction between the effects of social consensus and moral conviction on support for a polarized topic, as I previously theorized. Researchers’ inability to reliably manipulate social consensus and moral conviction continues to be a major impediment in synthesizing a cohesive framework for the effects of consensus and conviction (as well as their potential interaction) on polarized topics. However, these difficulties also indicate fruitful directions for future work; Developing more sophisticated tools to delineate between different characteristics of polarized topics will allow researchers to isolate promising targets for polarized belief change, and increased domain knowledge for what aspects of moral conviction are relevant for that specific population will ideally allow researchers a more thorough empirical test of the theorized relationship between social consensus and moral conviction.