# Conclusions, Interpretations, and Recommendations

The purpose of this dissertation was to understand and change polarized beliefs. To do this, we 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, a social consensus manipulation can change support for strongly held beliefs. We 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 (Is this true? If so in parentheses describe the factorial cross that yielded the four conditions). 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 our 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.

In light of the full set of results, some preliminary conclusions seem reasonable to draw. We were 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 our 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), we were 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 our topics were explicitly polarized (UHC, capital punishment, climate change, etc.).

Additionally, we were 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 the difficulty we had in successfully manipulating moral conviction. Thus, after developing further improvements in moral conviction manipulation, we could plausibly replicate this study and see significant 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 plausible 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 results we observed. We could investigate this further in future studies by presenting persuasive essays that are both in favor and opposition to the topic in question.

We were able to successfully manipulate perception of social consensus using a very simple experimental manipulation adapted from Kobayashi (2018). This bodes well for the generalizability of this method, as the intervention was successful even though 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 we used for this study explicitly relied on deceiving our participants as to the base rate of consensus for each of our 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 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. This would significantly increase the external validity of the conclusions that could be drawn. In a practical sense, many beliefs do not lend themselves to easily revealed preferences, so this suggestion would not be viable for all topics.

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 we used in our 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 isn’t 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, we find that manipulating social consensus seems to affect support for a variety of polarized topics, such that greater perception of social consensus in favor of a topic is associated with increased support, and the obverse is associated with decreased support. Furthermore, we found that we were 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, we 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 we previously theorized. Developing a better understanding of how to manipulate moral conviction is necessary to explore this potential interaction further.