# Study 3

**Method**

Study 3 analyzed the interaction between moral conviction and social consensus on polarized beliefs using a 2x2 within-subjects design. Participants were randomly assigned to one of two social consensus (low vs. high) and moral conviction manipulation conditions (moral responsibility vs. pragmatic framing). The primary outcome, support for [topic], was measured both before and after experimental manipulation. The Institutional Review Board at the University of Missouri reviewed and approved all submitted materials for Study 3.

***Participants***

A total of 491 undergraduate students 18 years of age or older at the University of Missouri participated in this study. Participants were recruited through an online survey platform and offered psychology course credit in exchange for their participation. Participants were asked to select categories that best described their race/ethnicity. Participants also self-selected their preferred gender identity.

***Materials and Procedure***

Participants began by reading our cover letter and providing consent. Next, they were provided with the Ethical Standards of Judgement Questionnaire. Then, for each of our three [topics] (universal health care, capital punishment, and usage of AI in the workplace), participants provided their initial level of support for the [topic] (our primary outcome), as well as how much moral conviction they have regarding their position. Additionally, our participants self-reported their level of familiarity with each [topic], as well as their openness to changing their mind. The three topics selected were designed to vary in general political orientation towards the topic, as well as the level of familiarity. Support for universal health care and capital punishment is traditionally associated with diametrically opposed political leanings (e.g., conservatives tend to support capital punishment and liberals tend to support universal health care), furthermore, both topics have been actively discussed in the US for decades, and thus Americans likely have reasonable familiarity with them (Stein, 2017; Bump, 2015). Usage of AI in the workplace was chosen as a topic explicitly because of its high level of novelty. The usage of AI is not currently seen in a politically polarized way, and the public does not have the same level of familiarity as the other two topics, given its relative newness (Fast & Horvitz, 2017).

Next, as in Study 2, participants were asked to read three short essays about universal health care, capital punishment, and the usage of AI in the workplace designed to manipulate the perception of moral conviction. They were randomized into one of two conditions: 1) Moral Responsibility or 2) Pragmatic. Thus, each participant received three essays, one on each topic, that all share the same moral framing. As in Study 2, We chose to focus on the moral responsibility and pragmatic framings because these conditions had the greatest between group differences in the preliminary data. All essays were readable at a high school level, as assessed by a Flesh-Kincaid readability score, and have comparable word counts.

Then, as in Study 1, participants estimated the proportion of the US population in 2018 that would be in support of the three issues. Afterwards, participants were given information about social consensus on both of these issues. To manipulate the perception of social consensus, participants were randomized into a ‘high social consensus’ or ‘low social consensus’ condition. In both conditions, participants were given feedback consisting of the base rate of support that the general American public (in 2018) had for the three issues. Participants in the ‘high social consensus’ condition saw results that were 20% higher than the true base rate. Participants in our ‘low social consensus’ condition saw results that were 20% lower than the true base rate. For example, if 65% of Americans agreed that capital punishment is necessary in the US, the high social consensus condition would be told that 85% agree, and the low social consensus condition would be told that 45% agree. After the social consensus information, participants were asked to indicate their degree of surprise at the stated level of public support and then estimate levels of public levels support in 2024. After receiving both the moral conviction and social consensus manipulations, participants again completed items measuring their level of support as well as their level of moral conviction for all three issues. Finally, participants completed several individual difference measures and provided demographic information.

***Measures***

**Primary Outcome.** Participant support for the issues was captured in the same way as Study 2, as a continuous variable ranging from strong disagreement (-50) to strong agreement (50) with the following statements: 1) “The US government needs to implement Universal Health Care because basic population needs are not being met.” (*Universal Healthcare*), 2) “Capital Punishment (the Death Penalty) is necessary in America” (*Capital Punishment*), and 3) “Americans should be able to use AI for job applications” (*Use of AI in the Workplace*). Likewise, moral conviction was assessed using the same composite measure as in Study 2.

**Secondary Outcomes**. Estimates of public support for the two issues were obtained by asking participants to estimate what percentage of the American public would agree with the above statements measuring support for [topic]. Participants provided a number ranging from 0-100%. Separate estimates were obtained for 2018 and 2024. Participants were also asked to rate how ‘surprised’ they were at the 2018 social consensus information provided. Surprise will be measured with a 5-point Likert scale ranging from ‘Not Surprised’ (1) to ‘Very Surprised’ (5). Topic familiarity was assessed by asking participants “How familiar are you with [topic]?”, measured as a continuous variable ranging from “I am extremely unfamiliar” (-50) to “I am extremely familiar” (50). Likewise, openness to belief change was assessed by asking participants “How open are you to changing your mind regarding your beliefs about [topic]?”, measured as a continuous variable ranging from “Extremely unlikely” (-50) to “Extremely likely” (50).

Individual differences in deontological and utilitarian orientation were measured using the Ethical Standards of Judgement Questionnaire (ESJQ) developed by Love, Salinas, and Rotman (2020). Six items measure deontological orientation (e.g., “Solutions to ethical problems are usually black and white”), and six items measure utilitarian orientation (e.g., “When people disagree over ethical matters, I strive for workable compromises”). Participant agreement with these statements will be measured with 5-point Likert scales ranging from ‘Strongly Disagree’ (1) to ‘Strongly Agree’ (5). Preliminary data from Study 2 indicated that each six-item subscale showed satisfactory internal consistencies with Cronbach’s α of .783 (deontology) and .750 (utilitarianism).

***Power and Statistical Analysis***

A minimum sample of 210 participants was needed to achieve 95% power for a 2x2 within-subjects ANOVA with two main effects and one 2-way interaction term. Power was determined a-priori using G-power 3.1.9.7 (Faul, Erdfelder, Lang, and Buchner, 2007; Faul, Erdfelder, Buchner, and Lang, 2009). Support for the three issues was treated as a continuous variable. We examined the effects of the moral conviction condition (increasing or decreasing moral conviction), the effect of the social consensus condition (high or low social consensus), as well as the interaction between moral conviction and social consensus on our outcome measure. All tests were conducted in R and considered statistically significant when P <.05.

***Study 3 Hypotheses***

Our first hypothesis predicted that high social consensus would lead to more positive support for issues (H1: a significant main effect of social consensus), which would be a replication of Study 1. Additionally, as our second hypothesis, we expect to observe a significant interaction between the social consensus and moral conviction manipulations on support for universal health care, capital punishment, and the usage of AI in the workplace (H2). Specifically, it is expected that increased moral conviction will reduce the effect of social consensus and conversely, decreased moral conviction will increase the effect of social consensus. We predicted this due to previous literature indicating that high levels of moral conviction inoculate individuals from the effects of social consensus; however, this has not been experimentally tested previously.

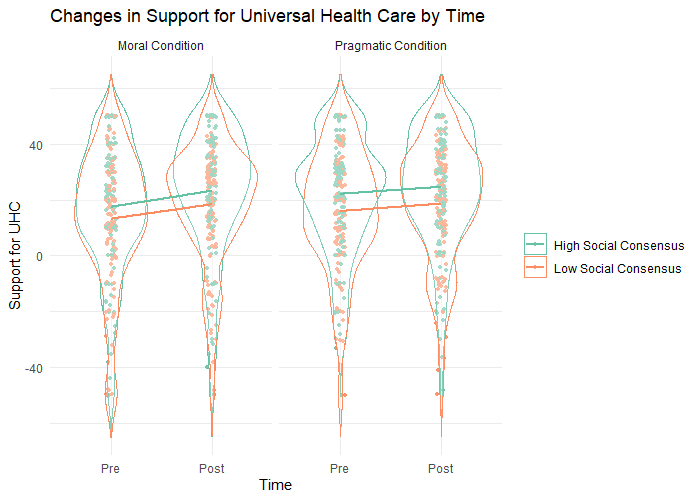
**Results**

We tested both hypotheses with a series of within-subjects analysis of variance (ANOVA) models comparing support for the issues both before and after our combined social consensus/moral conviction manipulation. The alpha level for these analyses was .05. Overwhelmingly, the pattern of results with regards to H1 and H2 held across all three [topics], thus in the interest of brevity, graphs indicating these relationships will be shown in the context of UHC only.

***Social Consensus and Moral Conviction Manipulation on Support for [Topic]***

Each of our three ANOVA models was composed of our dependent variable (quantified as the final level of support for our [topic], after both manipulations), with moral conviction manipulation condition, social consensus manipulation condition, initial (pre-manipulation) support for the [topic], initial [topic] familiarity, openness to belief change on [topic], and both utilitarian an deontological orientation as our ‘simple effect’ predictors. We also plan on examining the interaction of the moral conviction and social consensus manipulations. To test H1, we conducted an ANOVA model with our two manipulations (moral conviction and social consensus) as between-subjects factors.

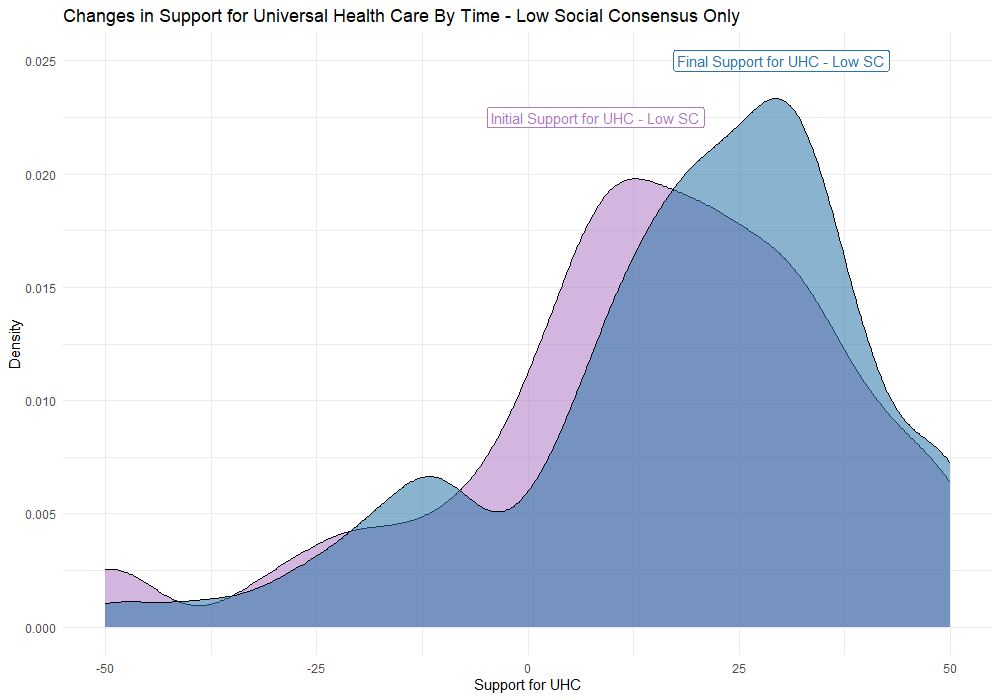
Our initial analysis showed no explicit support for H1. Our social consensus manipulations had no main effect on support for: 1) Universal Health Care, (ßconsensus condition = -1.712, *p* = *NS*); 2) Capital Punishment, (ßconsensus condition = -0.823, *p* = *NS*); or 3) AI in the Workplace, (ßconsensus condition = -2.67, *p* = *NS*).



Relatedly, our moral conviction manipulations likewise had no main effect on support for: 1) Universal Health Care, (ßconviction condition = -2.226, *p* = *NS*); 2) Capital Punishment, (ßconviction condition = -1.926, *p* = *NS*); or 3) AI in the Workplace, (ßconviction = -1.863, *p* = *NS*). Our initial analysis also showed no explicit support for H2. There did not seem to be a significant interaction between our social consensus manipulation, and our moral conviction manipulation for 1) Universal Health Care, (ßconsensus x conviction = -0.0762, *p* = *NS*); 2) Capital Punishment, (ßconsensus x conviction = -2.607, *p* = *NS*); or 3) AI in the Workplace, (ßconsensus x conviction = -3.31, *p* = *NS*). However, there was a significant main effect of openness to belief change on support for: 1) Universal Health Care, (ßchange = 0.0629, *p* = 0.05); 2) Capital Punishment, (ßchange = 0.121, *p* < .001); and 3) AI in the Workplace, (ßchange = 0.1745, *p* < .001). A graph of a number of dots

AI-generated content may be incorrect.

Furthermore, initial support for [topic] was significantly associated with final levels of support for 1) Universal Health Care, (ßinitial support = 0.715, *p* < .001); 2) Capital Punishment, (ßinitial support = 0.825, *p* < .001); and 3) AI in the Workplace, (ßinitial support = 0.680, *p* < .001). Additionally, utilitarian orientation was a significant predictor for support for the topic of ‘AI in the Workplace’ (ßutilitarian = 4.23, *p* = 0.029). Finally, a simple one sample t-test showed that the difference score comparing pre to post manipulation was not equal to zero, for 1) Universal Health Care, (*M* = 3.97, 95% CI [2.65, 5.29], *t(490)* = 5.92*, p <* .001); 2) Capital Punishment, (*M* = 6.16, 95% CI [4.54, 7.79], *t(490)* = 7.44*, p <* .001); and 3) AI in the Workplace, (*M* = 10.26, 95% CI [8.35, 12.17], *t(490)* = 10.54*, p <* .001). This shows that the intervention did increase support across all [topics].



A graph showing a line of a health care

AI-generated content may be incorrect.

***Moral Conviction Manipulation – Level of Moral Conviction Regarding [Topic]***

Expanding on our work from Study 2, we also wanted to see if the moral conviction manipulation successfully impacted perceived levels of moral conviction regarding [topic]. Each of our three ANOVA models was composed of our dependent variable (quantified as the final level of moral conviction for our [topic], after both manipulations), with moral conviction manipulation condition, social consensus manipulation condition, initial (pre-manipulation) moral conviction for the [topic], initial [topic] familiarity, openness to belief change on [topic], and both utilitarian an deontological orientation as our ‘simple effect’ predictors. We also planned on examining the interaction of the moral conviction and social consensus manipulations. A graph of a number of contours

AI-generated content may be incorrect.

Our initial analysis showed that our moral conviction manipulation was effective at changing perceived moral conviction for: 1) Universal Health Care, (ßconviction condition = -2.589, *p* = 0.038). However, our moral conviction manipulation was ineffective for the topics of: 1) Capital Punishment, (ßconviction condition = -1.215, *p* = *NS*); and 2) AI in the Workplace, (ßconviction condition = -2.084, *p* = *NS*). Unexpectedly, it seemed like our social consensus condition was associated with a change in perceived moral conviction for the topic of ‘AI in the Workplace’ (ßconviction condition = -3.426, *p* = 0.0185)

A graph of a graph showing a line of dots

AI-generated content may be incorrect.

Additionally, greater utilitarian orientation was associated with increased levels of final moral conviction for 1) Universal Health Care, (ßutilitarian = 3.659, *p* < .001); 2) Capital Punishment, (ßutilitarian = 2.896, *p* = .0045); and 3) AI in the Workplace, (ßutilitarian = 2.499, *p* = .0217). Conversely, greater deontological orientation was associated with decreased levels of final moral conviction for 1) Universal Health Care, (ßdeontological = -1.734, *p* = 0.033); and 2) Capital Punishment (ßdeontological = -2.259, *p* < .0094), but not AI in the Workplace, (ßdeontological = -1.569, *p* = .0964).

A graph of dots on a white background

AI-generated content may be incorrect.

Furthermore, initial moral conviction was a significant predictor of the final level of moral conviction, for 1) Universal Health Care, (ßinitial conviction = 0.683, *p* < .001); 2) Capital Punishment, (ßinitial conviction = 0.715 , *p* < .001); and 3) AI in the Workplace, (ßinitial conviction = 0.668, *p* < .001). Finally, the interaction between our moral conviction and social consensus manipulations was not significantly associated with our final levels of moral conviction for 1) Universal Health Care, (ßconsensus x conviction 0.4467 =, *p* = *NS*); 2) Capital Punishment, (ßconsensus x conviction = 0.957, *p* = *NS*); or 3) AI in the Workplace, (ßconsensus x conviction = 3.247, *p* = *NS*).

***Exploratory Analyses***

In addition to our planned analyses, we conducted additional exploratory analyses on baseline differences in moral conviction, openness to belief change, and familiarity regarding our three [topics]. For these analyses, we used a simple one-way ANOVA predicting moral conviction, openness to belief change, or topic familiarity, with topic (e.g., UHC, Climate Change, etc.) as our main predictor. One additional exploratory analysis we did was to determine if the effects of our paired manipulations were affected by the initial level of support for [topic]. We performed this exploratory analysis by first creating a categorical variable based on initial level of support, wherein participants that rated initial support from -50 to -5 were classified as ‘negative’, -5 to 5 were classified as ‘neutral’, and 5 to 50 were classified as ‘positive’. Then, our analysis was the same analysis as that done for H1, but with the addition of ‘initial level of support by category’ as an interaction with our original two-way interaction between social consensus and moral conviction condition (e.g., we are also looking at a potential three-way interaction). The table below indicates how many individuals were in each category, for each [topic].

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Negative** | **Neutral** | **Positive** |
| **UHC** | 72 | 47 | 372 |
| **Capital Punishment** | 260 | 51 | 180 |
| **AI in the Workplace** | 279 | 50 | 162 |

A diagram with red green and blue squares

AI-generated content may be incorrect.

Our first one-way ANOVA revealed that there was a statistically significant difference in openness to belief change between at least two of our topics (*F* (2, 1470) = 37.44, *p* < .001). A post hoc Tukey test showed that all three of our topics had significant differences in baseline openness to belief change at p < .05. Compared to our exploratory analysis from Study 2, we replicated the result that support for UHC is seen as significantly more open to belief change than the topic of capital punishment. Additionally, we see significant differences for openness to belief change regarding AI in the workplace, which we did not see for our other two topics in Study 2 (climate change and exercise).

A diagram with red green and blue squares

AI-generated content may be incorrect.

Our second one-way ANOVA revealed that there was a statistically significant difference in topic familiarity between at least two of our topics (*F* (2, 1470) = 132.7, *p* < .001). A post hoc Tukey test showed that all three of our topics had significant differences in topic familiarity at p < .05. We found that, contrary to *a-priori* expectations, our study sample self-reported the greatest familiarity with the topic of AI in the workplace, less familiarity with capital punishment, and even less with UHC. Qualitatively, the mean score for AI familiarity falls firmly in the range of ‘moderately’ familiar, which is unexpected, given the relative novelty of the field of AI as a whole (as compared to universal health care, or capital punishment, which has been in existence for decades).

A diagram with red green and blue squares

AI-generated content may be incorrect.

Our third one-way ANOVA revealed that there was a statistically significant difference in the initial level of moral conviction between at least two of our topics (*F* (2, 1470) = 62.12, *p* < .001). A post hoc Tukey test showed that all three of our topics had significant differences in baseline moral conviction at p < .05. As compared to our exploratory analysis in Study 2, none of our chosen topics were seen as lacking moral conviction (e.g., the exercise topic in Study 2 had a mean-value of 0 on the initial moral conviction scale). Additionally, we found that usage of AI in the workplace was held with the lowest, but still not nonzero amount of conviction, as compared to our other two topics.

A graph of a patient

AI-generated content may be incorrect.

Finally, our exploratory analysis on the effect of initial support for [topic] by category (e.g., negative, neutral, or positive) revealed that there was a significant two-way interaction between social consensus and initial categorically positive support (ßconsensus x positive support = -9.86, *p* = 0.0420) and a significant three-way interaction between conviction condition, social consensus, and categorically positive support (ßconviction x consensus x positive support = 14.18, *p* = 0.0463) for the topic of UHC, but not for capital punishment or usage of AI in the workplace. A sensitivity analysis indicated that, in theory, unobserved confounders explaining 7.71% of the residual variance for both the treatment and outcome would be sufficiently strong to explain away all the observed effect. Conversely, this indicates that unobserved confounders that do not explain at least 7.71% of the residual variance for both the treatment and the outcome would not be sufficiently strong enough to explain our observed effect.

***Discussion***

The results for Study 3 did not provide strong support for H1. We did not find a main effect of social consensus on support for [topic]. We also did not find strong support for H2, as we did not find a significant interaction between the effects of our social consensus and our moral conviction manipulations. However, we did find a significant difference in pre-post manipulation support, indicating that while there did not seem to be any difference regarding the effects of our intervention, there was a significant effect of the interventions taken as a whole, as compared to the initial level of support. We also found that openness to belief change and initial levels of support were strongly associated with the final level of support for [topic]. Additionally, our exploratory analysis indicated that there was a different pattern of support for individuals who started at a negative of neutral level of support for a [topic]. Primarily, the pattern indicated that the effect of the interventions on support is significantly greater for those starting at a negative or neutral level of support. Given that all interventions resulted in an increase in support, this is perhaps indicative of a ceiling effect.

Additionally, our moral conviction manipulation was only successful at actually changing perceived moral conviction in for the topic of UHC. This is indicative of either a significant difference in perspective on the topic of UHC as compared to capital punishment or AI, or that a different moral conviction manipulation may be more appropriate for further research on the matter. Regarding moral conviction, one extremely interesting finding is that greater utilitarian and deontological orientation were associated with moral conviction in opposite directions. Moral conviction can be defined as “the perception that one’s feelings about a given attitude object are based on one’s beliefs about right and wrong”. In comparisoin, deontological orientation is defined as ‘ethical rules clearly distinguish right from wrong’ and utilitarian orientation can be defined as ‘consequences are what distinguishes right from wrong’. Given that both moral axis (deontology and utilitarianism) are focused on what determines ‘right from wrong’, definitionally, greater orientation on both of these axis should plausibly be associated with greater moral conviction as a whole. However, we find in our study that greater deontological orientation is actually associated with decreased moral conviction, whereas greater utilitarian orientation is associated with increased moral conviction. Further research unpacking the mechanics behind this unexpected relationship is needed.