Does Implementing the Rationalization Principle Reduce Motivated Reasoning and Promote Belief Revision?

Paul Henne*, Matthew L. Stanley*, and Felipe De Brigard
Duke University

Author Note:
Paul Henne, Department of Philosophy, Duke University
Matthew L. Stanley, Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University
Felipe De Brigard, Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Department of Philosophy, Duke Institute for Brain Sciences, Duke University

*Henne and Stanley contributed equally to this project.

Correspondence should be addressed to:
Paul Henne
201 West Duke Building
Duke University
Durham, NC 27708
Paul.Henne@duke.edu
203-598-6890
Abstract: People frequently engage in biased, motivated reasoning about moral, social, and political issues, stymieing open, fact-based discussion. In two studies, this registered report aims to systemically implement a philosophical tool, the rationalization principle, so as to attempt to reduce motivated reasoning. In order to encourage the use of the strategy prescribed by the principle, we will have participants consider how a rational person with an opposing view might evaluate reasons for their view. We will investigate whether this manipulation moderates motivated reasoning and increases the likelihood that people change their minds about moral and socio-political issues.

Keywords: Moral Reasoning, Motivated Reasoning, Experimental Philosophy
After a lifetime of active membership and ardent support, Megan Phelps-Roper decided to leave the Westboro Baptist Church for good. A series of engagements on Twitter prompted her to consider why others disagreed with many of her moral, social, and political positions. In the midst of these engagements, Phelps-Roper realized that she needed “to listen” to her opponents because “we routinely refuse to acknowledge the flaws in our position or the merits in our opponent’s.” Eventually, she said publicly, “These shifts in my perspective contributed to a larger erosion of trust in my church, and eventually it made it impossible for me to stay”1. There are occasional anecdotes about individuals like Phelps-Roper changing their minds about moral, social, and political issues after shifting their perspectives when considering opponents’ reasons. And reasons, reasoning, and deliberation contribute to revising beliefs under a relatively narrow set of circumstances2-5. Nonetheless, a growing body of research suggests that people are generally resistant to changing their minds about moral and socio-political issues—even after considering compelling reasons and arguments that oppose their views6-10. And people frequently engage in biased, motivated reasoning to maintain their views. So, how can we promote less biased reasoning about contentious moral, social, and political issues in order to make belief revision more likely to occur? We will investigate whether a philosophical tool, the rationalization principle, can be systematically implemented to reduce motivated reasoning and increase the likelihood of decision change11.

It is now well-documented that people tend not to waver in their moral, social, and political positions, even in the face of compelling reasons and arguments for alternative positions9,10,12,13. When considering whether to, say, support the use of the death penalty for particularly heinous crimes, people may decide to support or to oppose it. Biased reasoning then helps them to maintain their initial decisions13,14,15,16,17. They might seek out new evidence that supports their decision (i.e., a confirmation bias)7,18, spend more time and effort denigrating evidence that does not support their decision (i.e., disconfirmation bias)6,19,20 or judge new reasons and evidence favoring their decision to be more compelling than reasons and evidence incongruent with their decision (i.e., a prior belief bias)6,7,9,18.
Although the current evidence is mixed\textsuperscript{21}, these biases sometimes even encourage people to become more entrenched in their existing decisions and beliefs (i.e., a backfire effect)\textsuperscript{22,23}. These systematic biases make it particularly challenging for people to change their minds and to engage in more sincere, fact-based, and open discussion about moral, social, and political issues.

These systematic biases in reasoning support \textit{directional goals}\textsuperscript{9,15,24} which motivate people to seek out and to evaluate evidence in a biased way in order to support their existing beliefs\textsuperscript{14}. Directional goals contrast with \textit{accuracy goals}, which motivate people to seek out and evenhandedly consider relevant evidence in order to reach a correct or optimal conclusion\textsuperscript{25-26}. When people pursue accuracy goals, they tend to invest cognitive effort into carefully assessing the relevant information while setting aside beliefs that could bias their evaluation\textsuperscript{14,27}. If people are likely to pursue directional goals at the expense of accuracy goals when reasoning about social, political, and moral issues, it is an open question as to how people can be encouraged to seek out and to evaluate reasons and evidence in an unbiased—or, at least, less biased—way.

There is some motivation to believe that perspective taking—\textit{i.e.}, the intentional mental simulation of another’s cognitive and emotional states (e.g., beliefs, intentions, or priorities)\textsuperscript{28}—may reduce directional reasoning when participants consider reasons opposing their initial decisions and positions. This perspective-taking strategy is used often in moral philosophy to evaluate premises in moral arguments, and, in order to shift people’s beliefs and intuitions, many philosophers encourage this practice of perspective taking in their thought experiments\textsuperscript{29}. Some developmental psychologists, moreover, argue that the capacity to change perspectives when thinking about moral issues is essential to late-stage development of moral reasoning and makes people more likely to revise and update their moral beliefs after reasoning about them\textsuperscript{30-31}. As this capacity for perspective taking declines with age so does higher-order moral reasoning\textsuperscript{32}. Perspective taking, furthermore, mitigates racial biases, produces more generous and forgiving explanations for negative outcomes experienced by others, and induces some positive behavioral changes\textsuperscript{32-37}. Perspective taking also seems to improve some kinds of analytic
reasoning—although the effects depend on the relationship to the participant and the strength of the opposing position\(^6\).\(^{31}\).\(^{39}\) It is, nonetheless, unclear if these changes in beliefs and decisions occur by making people more responsive to reasons and arguments that counter their initial position. Moreover, no empirical research shows perspective taking making a difference to moral decision making and morally-laden socio-political decision making.

We will investigate the utility of a novel philosophical strategy for reducing directional reasoning about moral and socio-political issues. While this strategy requires perspective taking, it also extends beyond the perspective-taking manipulations that have been implemented in other studies and potentially explains why so few studies show perspective-taking manipulations affecting moral decision change.

When interpreting someone’s position or argument, philosophers do not only take the perspective of their opponent and then consider the opponent’s arguments; rather, they use some principles to guide their consideration of opposing arguments. Related to the famous principle of charity\(^40\)-\(^42\)—\(i.e.,\) that one ought to represent another’s argument in the strongest possible way—is the rationalization principle: when interpreting someone’s beliefs and arguments, we should represent this other person as a rational agent and assume that this other person has good reasons for their beliefs\(^11\). While this principle is encouraged in philosophical practice in order to engender agreement, it remains an open question whether the strategy it prescribes actually reduces directional reasoning and promotes belief change about moral and socio-political issues outside of the philosophy classroom.

In two studies, we will ask participants to consider why someone who would make the opposing decision might find reasons for their view compelling. That is, we encourage participants to adopt the strategy prescribed by the principle of rationalization: to consider why an opponent might find reasons for their view compelling. By implementing this principle, we hope to combat biases in the reasoning process that encourage the pursuit of directional goals at the expense of accuracy goals. Using a variety of philosophical and everyday moral dilemmas adapted from previous work\(^9\)-\(^10\), we will first investigate whether using the strategy encouraged by the rationalization principle affects the likelihood that people
change their moral decisions (Study 1; $N_{expected} = 5,022$). We will then investigate whether this manipulation increases the likelihood that people change their minds about contemporary socio-political issues (Study 2; $N_{expected} = 5,022$). For moral dilemmas and socio-political issues, we will also investigate whether this strategy mitigates the biased evaluation of reasons for the opposing view (i.e., reduces the prior-belief bias). Reflecting on her opponents’ reasons and arguments from their perspective and seeing them as rational—rather than simply reflecting on their reasons—persuaded Phelps-Roper to change her beliefs about the Westboro Baptist Church and ultimately leave the organization for good. If this strategy can be systematically generalized, then it might encourage more sincere, fact-based, and open discussion about moral, social, and political issues.

Predictions

Participants will be randomly assigned to one of three between-subjects conditions in each study. Some participants will be encouraged to consider why someone with the opposing view might find reasons for their view compelling (rationalization condition), and other participants will be encouraged to consider why someone with the same view as the participant might not find reasons for the opposing view compelling (no-rationalization condition). We will also have a control condition (control condition) where participants will evaluate reasons without considering why another person might find the reasons compelling or not. We predict that motivated reasoning will decrease, belief change will increase, and confidence will decrease for participants who are randomly assigned to the rationalization condition relative to the no-rationalization condition and the control condition.

Study 1

Materials and Methods
Participants. Participants will be recruited from Amazon’s Mechanical Turk (AMT). Participant recruitment will be restricted to individuals in the United States who have at least 50 accepted HITs and a prior approval rating above 90%. Participants will be paid $.80 for their time.

Sample Size. Our primary analysis of interest is a Person’s chi-square test, which will be computed for each of the six dilemmas separately. With an alpha of .05, we aim to get a sample at .95 power.\textsuperscript{10} The effect size was calculated to be small-medium effect of .14. Hence, 761 participants are needed for each vignette (see Appendix A for R script for power analysis). We expect roughly 10% participants to be excluded for each vignette, so we will aim to collect data from 837 participants for each vignette. In total, we will stop collecting data when we have a total of 5,022 participants across a total of six moral dilemmas.

Exclusion Criteria. Participants who fail to answer all questions in the survey will be excluded, but not if they only fail to answer demographics questions. Participants who fail the explicit attention check at the end of the survey will be excluded from the analysis. For this attention check question, participants will be asked the following: do you feel that you paid attention, avoided distractions, and took the survey seriously? They responded by selecting one of the following: (1) no, I was distracted; (2) no, I had trouble paying attention; (3) no, I didn’t not take the study seriously; (4) no, something else affected my participation negatively; or (5) yes. Participants will be ensured that their responses would not affect their payment or their eligibility for future studies. Only those participants who select ‘5’ will be included in the analyses. No other exclusion criteria will be used. If we exclude more than 10% of our sample, we will recruit the number of participants over the 10% exclusion rate who were excluded from the study.

Materials. Six different two-option moral dilemmas have been adapted from previous work\textsuperscript{10} and will be used in this study (see Appendix B). Three of these dilemmas are adapted from philosophical work, and the other three describe more everyday moral situations. Taken together, then, these dilemmas offer diverse kinds of moral situations in order to increase the generalizability of our results. Each dilemma is accompanied by six different reasons for favoring each of the two possible decisions that were previously
crowd-sourced and normed. So, there are 12 reasons total accompanying each dilemma. The Qualtrics survey for a single dilemma that will be used in this study is attached to this preregistration (Appendix C).

Procedure. This study will be self-paced and consists of a single session. We will implement a 6 (dilemma: crying baby, concentration camp, transplant, car accident, cheating, or extra change) x 3 (condition: no rationalization, rationalization, or control) between-participants design. After providing informed consent, participants will be randomly assigned to carefully read one of the six different dilemmas. After reading the dilemma, participants will be instructed to make an initial decision indicating which of the two options they would choose in the dilemma. Participants then will be instructed to indicate whether they based their decision on (1) an immediate reaction or feeling after reading the dilemma such that no reasons affected the decision, (2) at least one reason, or (3) some other method. Participants will also indicate how confident they are that their chosen option is the better option in the dilemma (1 = not at all confident, 9 = very confident).

Next, participants will be randomly assigned to exactly one of three between-subjects conditions: no rationalization, rationalization, or control. Regardless of the condition, participants will only evaluate a set of reasons that conflict with the initial decision they made in the dilemma. That is, participants will be asked to evaluate the quality of the reasons for selecting the option they did not choose in the dilemma on a scale from 1 (worst possible reason) to 9 (best possible reason). These reasons will be presented in a random order on the screen, one at a time. Participants will also indicate whether they had considered each reason prior to seeing it in the study (binary response: yes, no). After they make their initial decision and before they evaluate the reasons, participants will be randomly assigned to a condition. In the no-rationalization condition, participants will be instructed to think about someone they know personally who might choose the same option in the dilemma. Then, they will be asked to think about and describe why this other person might think that each reason is not compelling. In the rationalization condition, participants will be instructed to think about someone they know personally who might choose the other
option in the dilemma. Then, they will be asked to think about and describe why this other person might think that each reason is compelling. In the control condition, participants will evaluate the quality of the reasons without imagining how any other person might evaluate the reasons. Participants assigned to the no-rationalization and rationalization conditions will be tested on their comprehension of the instructions, and they will receive feedback on their responses to the test question.

Regardless of the condition, participants will then be asked to make a final decision about what they would do in the dilemma, and they will rate how confident they are that their chosen decision is the better option of the two (1 = not at all confident, 9 = very confident). At this stage, participants have the opportunity to change their minds. Participants will be ensured that their response here will not affect whether we accept the HIT or their ability to participate in future studies. They then will answer several demographics questions (age, gender, race, education) and the explicit attention check question (see above). Upon completion, participants will be monetarily compensated for their time.

Analysis Plan

The $\alpha$ for all statistical tests will be set at 0.05.

Rationalization and decision change. First, we will investigate whether participants are more likely to change their initial decisions after considering reasons for the opposing positions in the rationalization condition relative to the other two conditions. We will compute new variables indicating whether participants changed or did not change their decisions after considering reasons (‘0’ = no change, ‘1’ = changed). We will report the percentage of participants who changed their initial decision for each condition collapsed across all dilemmas and for each individual dilemma. We will then determine whether there is a statistically significant relationship between condition and the percentage of participants changing their decisions versus not change their decisions. To that end, we will collapse across all six dilemmas, and use a $\chi^2$ tests to determine if there is statistically significant relationship between condition and the likelihood of decision change. We will also perform $\chi^2$ analyses between each
condition for each individual dilemma. For each analysis, we will report the \( p \)-value and the effect size (Cramer’s V). For any analysis here, if there are fewer than 5 participants who change their initial decision, a Fisher’s Exact Test will be performed. If such a test is used, we will report the \( p \)-value and Goodman and Kruskal \( \tau \). If we identify any significant effects, we will then use ransacking (Goodman 1969) between all possible pairs to determine which conditions differ.

Motivated evaluation of reasons. Second, we will investigate whether the biased, motivated evaluation of reasons for the opposing decision in the dilemmas is reduced in the rationalization condition relative to those participants’ evaluations in the no-rationalization condition and the control condition. Specifically, we will investigate whether participants in the rationalization condition evaluate reasons supporting the opposing decision in the dilemma more favorably than those participants in the other two conditions. To this end, we will initially compute a two-way ANOVA for each dilemma. For each ANOVA, the condition (no rationalization, rationalization, or control) and the initial decision (and their interaction) will be modeled as independent variables, and the dependent variable for each ANOVA will be the average rated quality of reasons evaluated. For each analysis, we will report the \( F \)-values, \( p \)-values, and the effect sizes (\( \eta^2 \)). If necessary, post-hoc tests between all possible pair-wise comparisons will be conducted using Tukey HSD, and \( t \)-values, \( p \)-values, and 95% CIs would be reported. Then, we will expand these models within a multiple regression framework to statistically control for the reported strategy used to make the initial decision and the number of novel reasons evaluated. In doing this, we expect that the effect of condition on the average rated quality of the reasons will remain statistically significant even after accounting for these additional variables. We will compute one multiple regression model for each dilemma, and we will report beta-values (standardized and unstandardized, \( t \)-statistics, \( p \)-values, and 95% CIs around beta-values.

Changes in decision confidence. We will then investigate whether there is a change in decision confidence from the initial decision to the final decision that differs as a function of the condition (no rationalization, rationalization, or control). In doing so, we will compute a new variable representing the
change in decision confidence from the initial decision to the final decision made after evaluating reasons (final decision confidence minus initial decision confidence). We will initially compute a one-way ANOVA for each dilemma. For each ANOVA, the condition will be the independent variable, and the dependent variable for each ANOVA will be the change in decision confidence. For each analysis, we will report the $F$-values, $p$-values, and effect sizes ($\eta^2$). If necessary, post-hoc tests between all possible pair-wise comparisons will be conducted using Tukey HSD, and $t$-values, $p$-values, and 95% CIs would be reported.

Study 2

Materials and Methods

Participants. Participants will be recruited from AMT. Participant recruitment will be restricted to individuals in the United States who have at least 50 accepted HITs and a prior approval rating above 90%. Participants will be paid $.80 for their time.

Sample Size. As in Study 1, our primary analysis of interest is a Person’s chi-square test. This statistical test will be computed for each of the three different socio-political issues separately. With an alpha of .05, we aimed to obtain a sample at .95 power. The effect size was calculated to be small-medium effect of .14. Hence, 761 participants are needed for each issue (see Appendix A for R script for power analysis). We expect roughly 10% participants to be excluded for each issue, so we will aim to collect data from 837 participants for each issue. In total, we will stop collecting data when we have a total of 5,022 participants across a total of six socio-political issues.

Exclusion Criteria. Participants who fail to answer all questions in the survey will be excluded, but not if they only fail to answer demographics questions. Participants who fail the explicit attention check at the end of the survey will be excluded from the analysis. For this attention check question, participants will be asked the following: do you feel that you paid attention, avoided distractions, and took the survey
seriously? They responded by selecting one of the following: (1) no, I was distracted; (2) no, I had trouble paying attention; (3) no, I didn’t not take the study seriously; (4) no, something else affected my participation negatively; or (5) yes. Participants will be ensured that their responses would not affect their payment or their eligibility for future studies. Only those participants who select ‘5’ will be included in the analyses. No other exclusion criteria will be used. If we exclude more than 10% of our sample, we will recruit the number of participants over the 10% exclusion rate who were excluded from the study to obtain achieve our desired sample size.

Materials. Three different two-option socio-political issues have been adapted from previous work and will be used in this study (see Appendix D). Each socio-political issue is accompanied by seven different reasons to support the presented position and seven different reasons to oppose the presented position. So, there are 14 reasons total accompanying each issue. The Qualtrics survey for a single issue that will be used in this study is attached this preregistration (Appendix E).

Procedure. This study will be self-paced and consist of a single session. We will implement a 3 (issue: animals in scientific testing, drone strikes, physician-assisted suicide, voter ID laws, fracking, affirmative action) x 3 (condition: no rationalization, rationalization, or control) between-participants design. After providing informed consent, participants will be randomly assigned to carefully read neutral, descriptive background information about one of the three socio-political issues. Then, participants will be asked about their prior knowledge regarding the issue. For example, participants will answer the following question for drone strikes: “Prior to reading the descriptive information above, how much did you personally know about United States drone strikes on military targets overseas?” (1 = nothing at all, 9 = very much). Participants will then be instructed to make an initial choice to either support or oppose the stated position. For example, for this particular issue, participants could choose between the following: (1) I support the use of drones by the United States to kill military targets overseas, or (2) I oppose the use of drones by the United States to kill military targets overseas. Then, they provided a judgment about their confidence in the superiority of their initial chosen position (1 = not at all confident; 9 = very
confident). Participants will then be instructed to indicate how they made their decision: (1) an immediate reaction or feeling after reading about the issue such that no reasons affected the decision, (2) at least one reason, or (3) some other strategy.

Next, participants will be randomly assigned to exactly one of three between-subjects conditions: no rationalization, rationalization, or control. Regardless of the condition, participants will only evaluate the set of reasons that conflict with the initial decision they made about the socio-political issue. That is, participants will be asked to evaluate the quality of the reasons for selecting the unchosen option on a scale from 1 (worst possible reason) to 9 (best possible reason). These will be presented randomly on the screen, one at a time. Participants will also indicate whether they had considered each reason prior to seeing it in the study (yes, no). In the no-rationalization condition, participants will be instructed to think about someone they know personally who might choose the same option. Then, they will be asked to think about and describe why this other person might think that each reason is poor. In the rationalization condition, participants will be instructed to think about someone they know personally who might choose the other option. Then, they will be asked to think about and describe why this other person might think that each reason is compelling. In the control condition, participants will just evaluate the quality of the reasons without imagining how any other person might evaluate the reasons. Participants assigned to the no-rationalization and rationalization conditions will be tested on their comprehension of the instructions, and they will receive feedback on their responses to the test question.

After providing all ratings for all reasons, participants will be asked to choose again their preferred option (i.e., to support or oppose the presented position). At this stage, participants have the opportunity to change their minds. They then provided a judgment about how confident they were that this final selected position is the superior position (1 = not at all confident; 9 = very confident). Finally, participants were asked if they considered any additional reasons not presented in the study. If participants considered other reasons, they were instructed to describe those additional reasons. They then will answer several
demographics questions and the explicit attention check question (see above). Upon completion, participants will be monetarily compensated for their time.

Analysis Plan

The $\alpha$ for all statistical tests will be set at 0.05.

*Rationalization and decision change.* First, we will investigate whether participants are more likely to change their initial decisions after considering reasons for the opposing decisions in the rationalization condition relative to the other two conditions. We will compute new variables indicating whether participants changed or did not change their decisions after considering reasons (‘0’ = no change, ‘1’ = changed). We will report the percentage of participants who changed their initial decision for each condition collapsed across all issues and for each individual issue. We will then determine whether there is a statistically significant relationship between condition and the percentage of participants changing their decisions versus not change their decisions. To that end, we will collapse across all six dilemmas, and use a $\chi^2$ tests to determine if there is statistically significant relationship between condition and the likelihood of decision change. We will also perform $\chi^2$ analyses between each condition for each individual issue. For each analysis, we will report the $p$-value and the effect size (Cramer’s V). For any analysis here, if there are fewer than 5 participants who change their initial decision, a Fisher’s Exact Test will be performed. If such a test is used, we will report the $p$-value and Goodman and Kruskal $\tau$. If we identify any significant effects, we will then use ransacking (Goodman 1969) between all possible pairs to determine which conditions differ.

*Motivated evaluation of reasons.* Second, we will investigate whether the biased, motivated evaluation of reasons for the opposing decision in the issues is reduced in the rationalization condition relative to those participants’ evaluations in the no-rationalization condition and the control condition. Specifically, we will investigate whether participants in the rationalization condition evaluate reasons supporting the opposing decision in the issue more favorably than those participants in the other two conditions. To this
end, we will initially compute a two-way ANOVA for each issue. For each ANOVA, the condition (no rationalization, rationalization, or control) and the initial decision (and their interaction) will be modeled as independent variables, and the dependent variable for each ANOVA will be the average rated quality of reasons evaluated. For each analysis, we will report the $F$-values, $p$-values, and the effect sizes ($\eta^2$). If necessary, post-hoc tests between all possible pair-wise comparisons will be conducted using Tukey HSD, and $t$-values, $p$-values, and 95% CIs would be reported. Then, we will expand these models within a multiple regression framework to statistically control for the reported strategy used to make the initial decision and the number of novel reasons evaluated. In doing this, we expect that the effect of condition on the average rated quality of the reasons will remain statistically significant even after accounting for these additional variables. We will compute one multiple regression model for each issue.

Changes in decision confidence. We will then investigate whether there is a change in decision confidence from the initial decision to the final decision that differs as a function of the condition (no rationalization, rationalization, or control). In doing so, we will compute a new variable representing the change in decision confidence from the initial decision to the final decision made after evaluating reasons (final decision confidence minus initial decision confidence). We will initially compute a one-way ANOVA for each issue. For each ANOVA, the condition will be the independent variable, and the dependent variable for each ANOVA will be the change in decision confidence. For each analysis, we will report the $F$-values, $p$-values, and effect sizes ($\eta^2$). If necessary, post-hoc tests between all possible pair-wise comparisons will be conducted using Tukey HSD, and $t$-values, $p$-values, and 95% CIs would be reported.

Acknowledgements

We thank Aleah Bowie, Aaron Ancell, Hannah Read, Gus Skorburg, and Kelly Shi for insightful comments and help with the manuscript. We also thank the IMC lab at Duke University for helpful comments and discussion about this project.
Author Contributions

PH and MS jointly conceived the study designs and analysis plans with input from FDB. PH and MS wrote the registered report with input and edits from FDB. PH and MS will collect and analyze the data. PH and MS will write the manuscript with input and edits from FDB.

Competing Interests

The authors declare no competing interests.

Data Availability

The datasets that will be generated and analyzed during the current study will be available in the OSF repository: [LINK].

Materials and Correspondence

Materials will be available at [LINK]. Additional requests should be addressed to PH.

References


