Chapter 5

Moral Perception

Introduction

During a July 2012 campaign stop, Barack Obama made some comments that he ended up regretting.

> "If you were successful, somebody along the line gave you some help. There was a great teacher somewhere in your life. Somebody helped to create this unbelievable American system that we have that allowed you to thrive. Somebody invested in roads and bridges. *If you've got a business – you didn't build that.* Somebody else made that happen." [1]

In their convention, the Republicans dedicated an entire day to the theme "We Built It," and President Obama's "you didn't build that" was featured ubiquitously in Republican advertisements.

This inartful construction by Obama tickled something in many conservative minds. It symbolized a president who, from their perspective, did not value hard work or individual effort. It represented a moral failing of a political party that champions the

[1] Emphasis added. From a transcript available at http://www.factcheck.org/2012/07/you-didnt-build-that-uncut-and-unedited/. There is some controversy about this line. It seems most likely to me that it resulted from Obama unintentionally skipping a line from the teleprompter. As with most campaign "gaffes," the statement sticks in the public conversation because it symbolizes some larger perceived or actual fault with the individual or political idea.

"takers" at the expense of the "makers." Conversely, many Democrats and liberals doubled down and saw the Republican response as a brash rejection of community, and a fundamentally hypocritical position given the fact that the Republican convention was being held at a venue that had been constructed mostly with government funds.[2] This deep moral divide was an important part of the overall story of the 2012 campaign.

Divides in politics often seem unbridgeable. As politics has polarized at the mass-level, politicians have seized the opportunity to exploit preexisting moral divides. In a certain sense this is inevitable and expected from election-minded politicians in a two-party system. Competition between the two parties should reveal these moral fault lines in the public (Kollman, Miller, and Page 1992, 1998), and once they are clear, they are too tempting to leave alone (Hillygus and Shields 2008).

In this chapter, I will examine the factors that lead individuals to react differently to moralized political rhetoric. I focus on two potential sources of individual-level variation in moral judgments. The first are stable individual-level moral predispositions. Differences in moral concern are measurable and exert a real impact on the ways in which individuals process political information as was shown in Chapter 3. The second factor I examine is the interaction between partisan cues and individual moral judgments.

Using data from ratings collected from 172 subjects who reacted to more than 1,000 sentences sampled from the *Congressional Record*, I show how these two factors – predispositions and partisan cues – affect moral judgments in significant and substantively important ways. My results suggest that moral judgments are domain specific. Individuals vary significantly in the emphasis they place on different moral domains, and these

[2]For example, http://mediamatters.org/blog/2012/08/22/fox-approved-convention-theme-contradicted-by-p/189507.

individual differences result in substantially different reactions to the same content. Individuals also react to partisan source cues in predictable ways. When statements are attributed to copartisans, they are received less critically than those that come from the other team. However, in many cases, the effects of partisan source cues are dwarfed by the effects of moral rhetoric that resonates with one's own moral priorities.

Unpacking the "Black Box"

Existing accounts of political cognition (for example, the extensive body of work collected by Lodge and Taber (2013)) focus on the *processes* by which individuals arrive at their political opinions. Lodge and Taber's experiments have convincingly and repeatedly demonstrated the importance of motivated reasoning in understanding how individuals process new information and incorporate prior information in their attitude reports. What is missing from these theories is an account of the motivations that feed into the original model. Moral foundations theory gives us one way to get at these motivations.

In broad strokes, the Lodge and Taber account shows that affect is a critical component of political evaluations. Immediately upon being exposed to some new piece of information, we experience a flash of positive or negative affect. This affective response colors subsequent processing and facilitates or inhibits other concepts. This process leads to reasoning that is systematically biased in favor of one's own predispositions.

In most of the work on information processing, individual predispositions (the factors that lead to positive or negative affect) are "black-boxed." For example, liberals are expected to experience negative affect when evaluating conservative issue positions and

the Republican Party, but conservatives are expected to have an opposite pattern of reactions. From this starting point, those who study information processing can make predictions and evaluate hypotheses about the consequences of motivated reasoning. In this chapter, I am interested in shifting the focus back to the factors that lead different people to react to the same stimuli in different ways in the first place.

One of the key aims of this chapter is to establish the validity of the pre-political dispositions in realistic political settings. Moral foundations theory describes stable individual differences in moral outlook. In this dissertation, I am proposing that these individual differences in *moral* concern are translated into meaningful *political* differences. If an individual's moral outlook colors his or her perception of political information, we will have a piece of the story as to how these stable individual predispositions translate into political divisions.

Differential emotional reactions

I expect that individual differences in moral concern (as measured by responses to the Moral Foundations Questionnaire) will explain some of the variance in moral judgments and the intensity of emotional reactions to political arguments. To paraphrase an old cliché, morality is in the eye of the beholder.[3]

The first aim of my analyses will be to show that individual differences in predispositions are related to differential emotional reactions to the same stimuli. These differential emotional responses should be domain specific. I expect different moral foundations to be salient in different contexts. For example, individual differences on the Care/Harm

[3]This is not to say that moral foundations theory necessarily implies moral relativism. One need not take a position on the existence of objective "moral facts" to observe that moral perceptions are definitionally subjective.

measure should be more important to evaluations of arguments that focus on the hurt experienced by victims than would be the case for arguments that appeal to undeserving treatment (where I would expect the Fairness/Cheating domain to be relatively more influential).

Emotional Content of Moral Evaluations

The second major aim of this chapter is to show how moral evaluations are connected to emotional reactions. Individuals who experience stronger (either negative or positive) reactions to a political statement are expected to be more likely to feel that it is connected to their fundamental ideas about right and wrong.

Recent theoretical and empirical work in moral psychology suggests that a large portion of our moral judgments are driven by automatic affective reactions. The same work that informs Lodge and Taber's model of political information processing serves as the foundation of the affective turn in moral psychology. In laboratory experiments, researchers have manipulated affective states and found significant effects on moral judgment (Schnall, Benton, and Harvey 2008; Valdesolo and DeSteno 2006; Eskine, Kacinik, and Prinz 2011). These findings are in line with an intuitionist conception of moral judgment (Haidt 2001*b*; Prinz 2006).[4] Consistent with this recent line of research, I expect that more intense emotional reactions will be related to an increased probability of making a moral judgment.

[4]This point is not settled in the psychological literature. Pizarro and Bloom (2003) argue that rational deliberation plays a much larger role in informing our intuitions than is generally allowed in Haidt's model. It has been difficult to settle this controversy with laboratory experiments.

Partisan bias

Finally, I will look at the interaction between party identification, partisan source cues, and moral judgment. I expect that receiving source cues will significantly alter moral judgments. Same-party cues are likely to magnify affect and thus lead to a higher likelihood of moral classification. Out-party cues are expected to have the opposite effect – dampening affective responses and leading to a decreased likelihood of moral classification.

This final part of the analysis examines how partisanship biases moral processing. In-party cues should be seen as more closely related to an individual's sense of right and wrong, while those from the out-party should show the reverse pattern. Partisan biases could occur through (at least) two avenues. First, affect associated with the party is transferred to the content of the sentence (consistent with an "affect transfer" from a liked or disliked group to the stimulus in question). Alternatively, it could be that individuals make inferences based on other information they have about the parties (consistent with the "reputational premium," e.g. Sniderman and Stiglitz (2012)). My data are insufficient to distinguish between these two accounts, but understanding how moral judgments are made in more realistic political settings where partisanship is (often) known will help to paint a more complete picture of how politics becomes moralized.

Data and Methods

Subject Recruitment

In order to test my claims about the differences in individual moral perception, I collected data from Amazon's Mechanical Turk. Mechanical Turk data has been used in a growing number of social science applications, and a great many findings from more traditional samples have been replicated using Mechanical Turk workers (Berinsky, Huber, and Lenz 2012; Paolacci and Chandler 2014).[5] My data were collected in January of 2014.

A group of "Turkers" filled out an initial screening questionnaire in exchange for a small monetary incentive as well as the opportunity to participate in the second part of the study and receive a larger monetary reward for satisfactory participation. In the screening questionnaire, individuals were asked their partisanship and several other questions tapping political attitudes. They also filled out a version of the Moral Foundations Questionnaire.[6] Turkers tend to lean left in their politics, and although I made an effort to contact an equal number of Democrats and Republicans, the final dataset includes substantially more Democratic raters than Republicans (114 and 63 respectively).

[5]There are many justifiable concerns about Mechanical Turk samples. Mechanical Turk is well-suited for the purposes of this study for several reasons. First, the task (rating a series of sentences) closely mirrors the kinds of tasks for which Mechanical Turk was originally designed (crowd-sourcing type "human intelligence tasks"). There are some concerns that Turkers are acclimated to economic games and other types of experimental manipulations. It should also be noted that, in many cases, results from Mechanical Turk samples closely mirror those obtained from other populations. However, my study does not employ any of these kinds of manipulations, and it is unlikely that participants in the task outlined below would have incentives to give untruthful answers or otherwise contaminate the data. One additional concern relates to the degree of "moral diversity" that exists among Turkers. In my own and others' experience, Turkers tend to be younger, more educated, and more liberal than the general population. In an effort to maximize the diversity of the Turker population, I made an effort to oversample Republican identifiers.

[6]I had individuals fill out a slightly modified version of the MFQ. See Appendix A for question wordings.

A week or so after completing the screening questionnaire, individuals who qualified[7] were invited to participate in the second part of the study where they were asked to rate sentences drawn from the *Congressional Record*.

Sampling Political Texts

The sentences were sampled from several different topic areas. I tried to balance three considerations when pulling the sample of sentences from the *Congressional Record*. First, I wanted an equal number of Republican and Democratic statements, second I wanted to compare statements that were roughly similar in their substantive content, and finally I wanted some variance in the emotional and moral content of the sampled sentences. Ultimately, I constructed a stratified sample from several main issue areas: Abortion, Commerce, Defense, Education, Health Care, Immigration, Taxes, and Terrorism.[8]

Furthermore, I oversampled sentences that contained morally charged words.[9] In each topic area, I oversampled sentences with moralized language from speeches made by

[7]Only Turkers who passed the attention checks in the screening survey were qualified to participate in the follow-up study. I also made some attempt to oversample Republicans in the screening survey through quotas that appear to have not worked (the survey was supposed to cap participation by Democrats and continue collecting responses from Republicans).

[8]The topic classifications were based on the section of the *Congressional Record* that the speeches were pulled from. The Abortion topic included sentences from any speech falling within a section of the *Congressional Record* with "abortion" in its title. The Defense topic included speeches made during the scheduled debate over the Department of Defense Appropriations Act on July 18, 2012. The Commerce topic included speeches made on May 8, 2012 during the debate over the Commerce, Justice, Science, and Related Agencies Appropriations Act.

[9]Morally charged words were identified from a slightly pared down version of the Moral Foundations Dictionary available at www.yourmorals.org. In an effort to minimize false-positives, I excluded several phrases from the data where a putatively "moral" word as defined by the Moral Foundations Dictionary is used in a procedural or parliamentary way. I excluded the phrase "balance of my time" as "balance" here is being used in a non-moral context. Other phrases excluded were, "health care", "point of order", "in order", "on the order", "executive order." I entirely excluded "law" and "submi*" from the dictionary as these words were most often used in non-moral ways.

Members of Congress from each party. I made an effort to strip out purely parliamentary and procedural language from the population of sentences prior to pulling the sample. Ultimately, I received usable ratings on 1,074 unique sentences. Sentences were presented to the Turkers such that some would be rated many times. The average sentence was rated almost 3 times, but many sentences were rated only once and a few sentences were rated five or more times.

Procedure

For the first set of sentences, the Turkers were informed that the sentences were taken from speeches made by members of Congress, but they were not given any additional information about the speakers' identities. For the last five sentences that they rated, they were provided with the partisanship of the speaker (e.g. "A Republican representative said: ..."). In all, the subjects recorded nearly 2,700 ratings.

Respondents were then asked several questions about each sentence:

a. How strong was your emotional response to reading the sentence? (7 point scale from "Very strong negative reaction" to "Very strong positive reaction" with a midpoint of "Neither positive nor negative reaction");

b. To what extent does this sentence connect to your own beliefs about fundamental questions of right and wrong? (5 point scale from "Not at all" to "A great deal")[10]

c. (Only for individuals who indicated that the sentence had some moral content) If you had to choose one from the list below, would you say this sentence refers to:

    Norms of harm or care (e.g. unkindness or kindness, causing pain to another or caring for someone in pain)

    Norms of fairness or justice (e.g. cheating or honesty, reducing equality or "leveling the playing field")

[10]This item was adapted from one of the items in Linda Skitka's moral conviction scale (Skitka, Bauman, and Sargis 2005; Skitka 2010).

Norms of loyalty (e.g. betrayal of a group or being a good team player)

Norms of respecting authority (e.g. subversion or insubordination, showing proper respect for authority

Norms of purity (e.g. degrading or disgusting acts, sanctified or ennobling acts)

Some other moral norm (please specify):

Results

Differential Moral Evalutaion

To test whether individual differences in the moral foundations measures are related to the subjects' ratings of the content of each sentence, I first fit a model to estimate the moral categorization of each sentence. These models produced estimates of the degree to which each sentence fit within a particular moral domain.

We can think of individual categorizations as composed of two parts. The first part is a function of the objective or unbiased content (α_j),[11] and the second part is a function of subjective elements unique to each subject ($\beta_i + \delta_{ij}$). This second part of the categorization can be broken down further into two elements. The first is stable to the individual subject (β_i) and the second is a function of the interaction between the subject and the text (δ_{ij}). This is expressed mathematically below in a straightforward analysis of variance model:

$$y_{ij} = \alpha_j + \beta_i + \delta_{ij}$$

The equation above naturally lends itself to a hierarchical framework where we consider each constituent part of the expression to be itself a function of higher level factors.

[11] It might be more appropriate to label this as the "consensus" component.

For example, the sentence-specific component, α_j, could be modeled as a function of the topic of the sentence, the partisan affiliation of the sentence's author, and other sentence-level covariates. The β_i could be written as a function of individual subject demographics that might be expected to affect their evaluations.

Following the framework outlined above, the specific moral category was modeled as a function of sentence specific-factors (the partisanship of the author, the issue area, a measurement of the "polarity" of the sentence based on the words used,[12] a set of variables derived from the Moral Foundations Dictionary that hold the counts of various morally charged words in the different moral domains) and a set of subject-specific factors (age, education, party, sex, moral foundations scores). The subject-sentence interaction term controlled for whether or not individuals were cued with the partisanship of the sentence's author (and specifically whether or not the partisanship of the author matched the partisanship of the subject).[13]

$$\alpha_j = a_{0j} + a_1 * AuthParty_j + a_2 * Issue_j + a_3 * Polarity_j + a_4 * Polarity_j^2 + \sum_{m=1}^{12} A_m * MoralWords_{jm}$$

$$\beta_i = b_0 + b_1 * Age_i + b_2 * Educ_i + b3 * SubjParty_i + \sum_{m=1}^{5} B_m * MF_{jm}$$

$$\delta_{ij} = d_1 * Cue_{ij} * I(AuthParty_j = SubjParty_i) + d_2 * Cue_{ij} * I(AuthParty_j \neq SubjParty_i)$$

$$Pr(y_{ij} = 1) = \Phi(\alpha_j + \beta_i + \delta_{ij})$$

[12] Polarity was calculated by processing the sampled sentences through the qdap package in R 3.0. The algorithm uses affectively charged words and their modifiers to calculate a valence score for each sentence with low values corresponding to more negative words and hight values corresponding to more positive words.

[13] All of the results reported below are robust to restricting the data to the first fifteen ratings where no source cue was given.

Individual subject classifications were fit with a multinomial distribution with six categories (the five moral domains an other/non-moral category). The model produced estimates of the probability that each sentence belonged to one of the six categories (with the probabilities summing to unity). The models were estimated in JAGS 3.0 (Plummer et al. 2003) with vague priors centered around zero for the hyperparameters. The model code can be found in Appendix D.

The full results for the models used to generate the sentence categorizations can be found in Appendix D. A few points from the estimation are worth noting. Individual differences on the moral foundations did not seem to systematically affect how the sentences were coded. Only one of the coefficients associated with the moral foundations of the individual coders was significantly different from zero. This means, for example, that coders who are particularly sensitive to violations of the Care/Harm foundation were no more likely to code a sentence as belonging to that domain than coders who place relatively less emphasis on that domain. At least when it comes to moral categorization, individuals' own moral outlooks do not seem to bias responses in a systematic way. The one exception to this was in the Loyalty domain. There was a significantly positive relationship between an individuals' score on the Loyalty foundation and the likelihood of classifying a sentence as belonging to the Loyalty domain.

At the sentence-level, the topic of the sentence is a reliably significant factor in determining the moral domain to which it belongs. For example, sentences that were drawn from speeches about abortion were more likely to be classified as belonging to the Care/Harm and Sanctity/Degradation categories. The specific moral words used in each sentence were also important. On this score, the Moral Foundations Dictionary seems to perform relatively well as a predictor of the appropriate moral domain. For the most

part, the coefficients associated with the different categories in the Moral Foundations Dictionary behave as we would expect. Sentences that contain the words associated with each foundation were generally significantly more likely to be classified as belonging to that domain.[14]

Using the α_j estimates from the model, we can identify sentences that belong to each domain. In the examples below, I have selected a few of the sentences with the highest baseline likelihood of being classified in a particular domain.

a. Care/Harm Domain:

"Tens of thousands suffer more from wounds both visible and invisible."

"As a daughter and wife of physicians, I am shocked that we would so quickly dismiss the judgment of our country's medical personnel and families in making the best decision to preserve the health and lives of their loved ones."

b. Fairness/Cheating Domain:

"We believe that our nation's disabled veterans and wounded warriors have waited long enough for access to pools and spas."

"Unfortunately, unfair trade practices from countries like China make this increasingly difficult."

c. Loyalty/Betrayal Domain

"They are at war with us by supporting and funding the very terrorists that we are up against."

"The Army, the National Guard, and the National Guard Association strongly oppose this amendment."

d. Respect/Subversion Domain

"It's about respect for the rule of law."

"It is time we did the jobs we swore allegiance to the Constitution to do, even if others will not."

"For those watching at home, what would happen to them if they ignored a summons for jury duty?"

[14]The Loyalty and Sanctity domain-specific words from the Moral Foundations Dictionary were least associated with subjects' classifications. This might be due in part to the relatively small amount of Loyalty and Sanctity language in the texts.

Party	Topic	n	Care	Fair	Loyal	Respect	Sanctity
R	Abortion	224	0.24	0.28	0.04	0.02	0.06
D	Abortion	222	0.23	0.36	0.03	0.02	0.03
R	Defense	216	0.13	0.26	0.12	0.07	0.00
D	Defense	222	0.17	0.32	0.1	0.06	0.00
R	Commerce	219	0.11	0.43	0.07	0.06	0.00
D	Commerce	200	0.1	0.5	0.06	0.06	0.01

Table 5.1: Mean classification (scaled from 0 to 1) for sentences by party and topic

e. Sanctity/Degradation Domain

"Without it, people could be forced to participate in something they strongly believe to be morally wrong."

"But I also believe it's morally wrong to take the taxpayer dollars of millions of pro-life Americans and use it to fund a procedure that they find morally offensive."

"In other words, the victim of rape had to show wounds and other matters that she really was forcibly raped before she could be covered, but they changed that because there was such an outcry."

Table 5.1 presents some descriptive statistics for the categorization measures. The parties were about evenly matched when it comes to the Care/Harm domain. Sentences taken from speeches by Democrats were slightly more likely to be classified as belonging to the Fairness domain than those from Republican speeches. Republicans had a slight edge when it comes to the binding foundations, but the differences between the parties are not very dramatic for this sample. The other striking thing to note about the classifications is the relative rarity of sentences belonging to the Loyalty, Respect, and especially Sanctity foundations.

Given these measurements of the domain-classification for each sentence, we can test hypotheses about differential moral perception among the of coders. If it is true that the moral foundations structure evaluations and moral judgment in significant ways, we would expect individuals who score highest on a particular foundation to report more

intense emotional reactions when reacting to sentences that are classified in the same domain. For example, an individual who places relatively more weight on the Care/Harm domain is expected to have a more intense emotional reaction to a statement relating to ideas about caring for the sick when compared against an individual who does not place as much emphasis on the Care/Harm domain.

To determine whether or not individuals are biased in their reactions to moral content based on their own moral outlook, I estimated a set of ordered probit regressions. The dependent variable was the subjects' self-reported intensity of emotional reaction to the content of the sentence. On the right-hand side of the regression equation, I included controls for the demographics of the rater (to capture systematic differences in emotional response) and some sentence-level controls to capture baseline differences in the content of the sentences. The key explanatory variables of interest are the main effect of the moral foundation of the rater, the main effect of the moral categorization of the sentence, and the interactive effect of the rater's score on the moral foundation and the sentence's categorization score.

The full regression results can be found in Appendix D. Figures 5.1 and 5.2 show the predicted probabilities generated from the model. Figure 5.1 shows the predicted probabilities of a rater reporting an extreme emotional reaction to the sentence (either "Very Negative" or "Very Positive"), and Figure 5.2 shows the predicted probability of giving a non-neutral emotional response.[15] The solid lines show how the probabilities change for individuals who place relatively more emphasis on a particular foundation (the

[15]Predicted probabilities were calculated by holding all other variables in the model at their mean values. I used a bootstrapping procedure to calculate the predicted probabilities as the delta method was giving nonsensical results (probabilities greater than 1 and less than 0). The models were estimated with clustered standard errors by rater to account for the nested structure of the data.

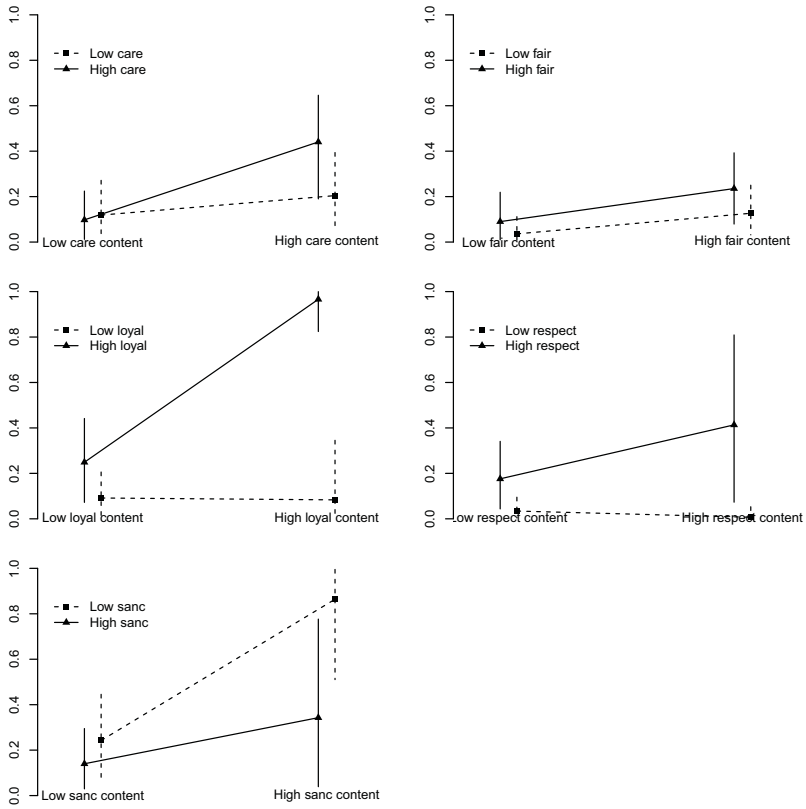

Figure 5.1: Emotional Reactions by Rater Moral Foundations and Topic Categorization. The plots show the predicted probability of giving the most extreme emotional response given the rater's moral foundation score and the categorization of the sentence content.

mean of the raters plus 1.5 standard deviations). The dashed lines show the predicted probabilities for individuals who place less emphasis on a particular foundation (mean minus 1.5 standard deviations).

For almost every foundation (excepting the Sanctity foundation), there is the expected positive interaction between the rater's score on the relevant moral foundation and the categorization of the sentence as belonging to that moral domain. The coders reported having more intense emotional reactions to the sentences that touched on particular domains of morality that they care about. The result is statistically significant

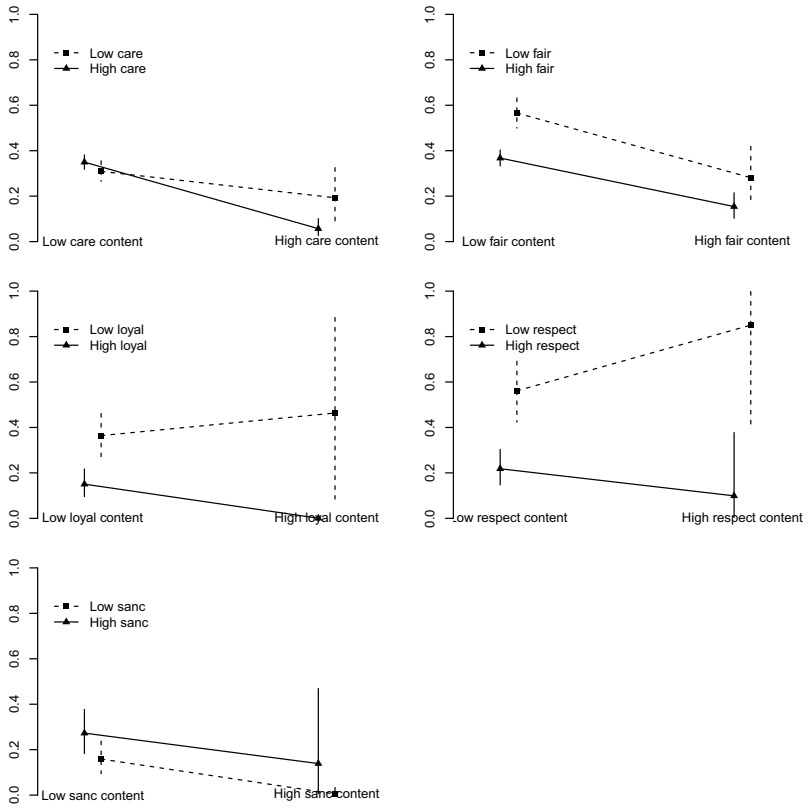

Figure 5.2: Emotional Reactions by Rater Moral Foundations and Topic Categorization. The plots show the predicted probability of giving a non-neutral emotional response given the rater's moral foundation score and the categorization of the sentence content.

for the Care, Loyalty and Respect foundations. For sentences belonging to the Fairness[16] foundation, there was no statistically significant difference between individuals who are high and low on the foundation and their emotional reaction to the content. The Sanctity/Degradation foundation actually showed the opposite pattern.[17]

As a test of the robustness of the above findings, we can run a series of placebo-like analyses. These tests involve looking for the "wrong" interaction. For example, we could look at the interaction between a sentence that was coded as belonging to the fairness foundation and the coder's score on the care foundation. The interaction between these mismatched foundations should not be statistically significant. When these alternative analyses are run almost none of the interactions emerge as significant. There were only two exceptions out of the twenty different tests performed. There was a statistically significant and negative interaction between sentences that were rated as belonging to the Respect domain and the Care/Harm moral foundation (e.g. individuals who scored highly on the Care/Harm foundation were significantly less like to register an emotional response to sentences coded as belonging to the Respect domain). Additionally, there was a statistically significant positive relationship between the sanctity foundation and sentences that were coded as belonging to the respect foundation.[18]

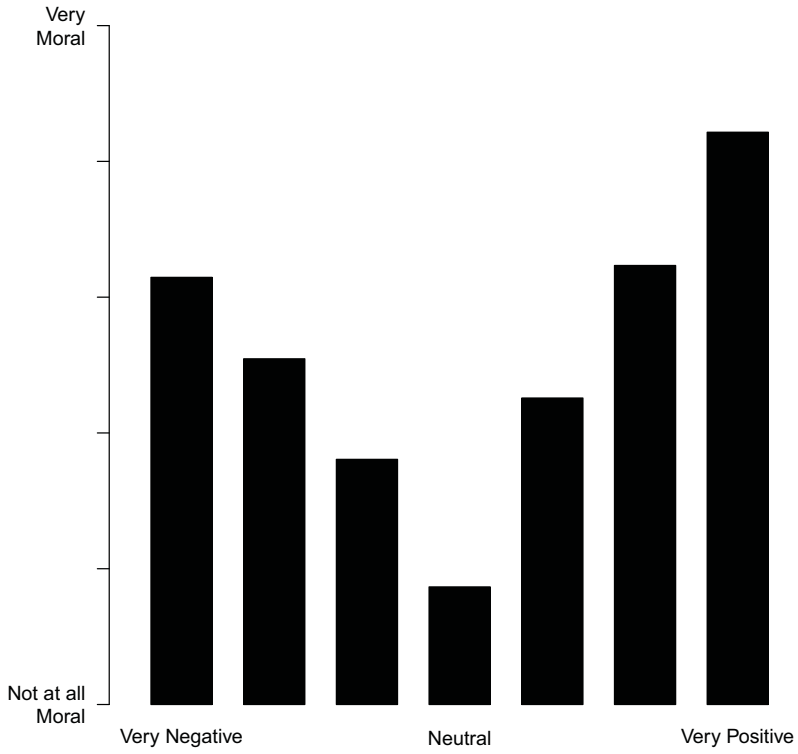

Figure 5.3: Moral Evaluation by Emotional Reaction. Bars show the average moral classifications for each scale point of the emotional reaction item.

The Emotional Content of Moral Evaluation

In line with the intuitionist model of moral judgement, there is a strong relationship between an individual's emotional reaction to a text and her moral classification. A simple bar graph reveals a strong relationship between an individual's emotional reaction to a particular sentence and his or her moral evaluation of that same text. Figure 5.3 shows the bivariate relationship.

This bivariate relationship holds up when controls for rater demographics (age, sex, education, partisanship) and characteristics of the sentences (topic, morally charged words, affectively laden words) are added to the model. The full results of the regression are available in Appendix D. In Figure 5.4, I show the predicted probability of registering the maximum moral evaluation (sentence connects to the raters "own fundamental ideas about right and wrong" "a great deal") across the range of emotional reactions (coded from "Very negative" = -1 to "Very positive" = 1).

The model allows for emotion to have an asymmetric effect, and indeed an asymmetry emerges. Intense positive reactions are more likely to lead to moral classifications than intense negative reactions. This difference is statistically significant, and substantively rather large. Raters who react "very negatively" to a sentence have a predicted 0.27 probability of rating that sentence as connecting "a great deal" to their own moral code. This compares with a 0.62 probability for raters who react "very positively." The 95 percent confidence interval for this difference ranges from a lower bound of 27 points to

[16]Further evidence that there might be something wrong with the way that the Fairness/Cheating dimension is being operationalized.

[17]This may be due to the rarity of language that was classified as belonging to the Sanctity foundation.

[18]This relationship is not entirely unexpected due to the high level of conceptual overlap between these two foundations. Indeed, I worried that these placebo tests would produce more statistically significant relationships especially between the Care and Fairness foundations and the three binding foundations.

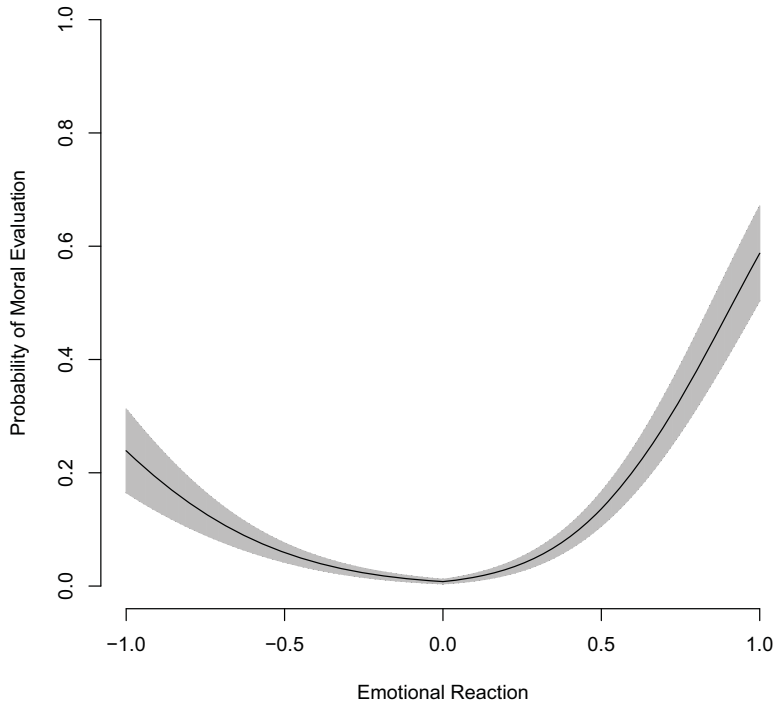

Figure 5.4: Moral Evaluation by Emotional Reaction. Predicted probability of maximum moral evaluation across the range of emotional reactions. The solid line shows the predicted value and the shaded region shows the 95% confidence interval.

an upper bound of 45 points.[19]

As expected, emotional reactions are strongly related to moral judgments. The asymmetry in positive and negative affect is perhaps a little surprising. Especially given what we know about loss aversion and the general primacy of negative affect, we might have expected the asymmetry to run in the other direction with negative affect being a more powerful moral motivator. One potential explanation for this counterintuitive asymmetry might be found in the wording of the moral judgment item. Respondents were asked, "To what extent does this sentence connect to your own beliefs about fundamental questions of right and wrong?" It is possible that some respondents interpreted this item as pertaining only to statements that they agreed with. Future research along these lines might consider prefacing this item with something like, "Regardless of whether you personally agree or disagree with the statement, ..."

Partisan Bias in Moral Evaluations

My third hypothesis relates to the ways in which individuals incorporate partisan cues into their moral evaluations. For the last five sentences presented to each subject, I provided the partisanship of the statement's author. This allows us to investigate how partisan biases might color moral and emotional reactions to the content.

One way in which this bias might manifest itself is in the emotional reactions to the text itself. It could be that, when presented with an in-party cue, partisans will have a stronger reaction than when presented with an out-party cue or no cue. This is straightforward to test with regression, and it does turn out to be the case. Holding

[19]The difference in predicted probabilities was generated through a bootstrapped simulation holding all other variables at their mean values.

other factors[20] constant at their mean values, an in-party cue makes it about 4 percentage points more likely that a rater will have a "very strong" emotional reaction and about 8 percentage points less likely that a rater will report having no emotional reaction. This effect isn't particularly large, but it is striking that the mere mention of partisanship has a robust effect on emotional reactions. There is no corresponding effect for an out-party cue.

We could similarly examine the effect of the party cue on moral evaluations. Above, I showed that emotional reactions were strongly related to moral evaluations, so we would expect to find an at least indirect effect of partisan source cues on moral evaluations operating through the demonstrably positive effect of in-party cues on emotional reactions. It also may be the case that there is a direct effect of party source cues on moral evaluations.

Fitting a regression of party source cues on moral evaluations leads to the opposite pattern of effects as we saw in the previous analysis. Whereas same party cues seemed to cause more extreme emotional reactions, it is out-party cues that do all of the work in this model. Receiving an out-party cue, all else constant, was associated with a drop in the probability of saying that a particular sentence was "quite a bit" or "a great deal" related to one's basic sense of morality. Unlike the emotional reactions model, the effect of partisan cues on moral evaluations seems to be concentrated among Republican identifiers.

[20]The, by now, standard set of rater-level (age, education, sex, partisanship) and sentence-level (topic, moral words, polarity) controls. Full regression results are available in the appendix. The results presented were fit with clustered standard errors at the rater-level. The findings are robust to including sentence-level fixed effects.

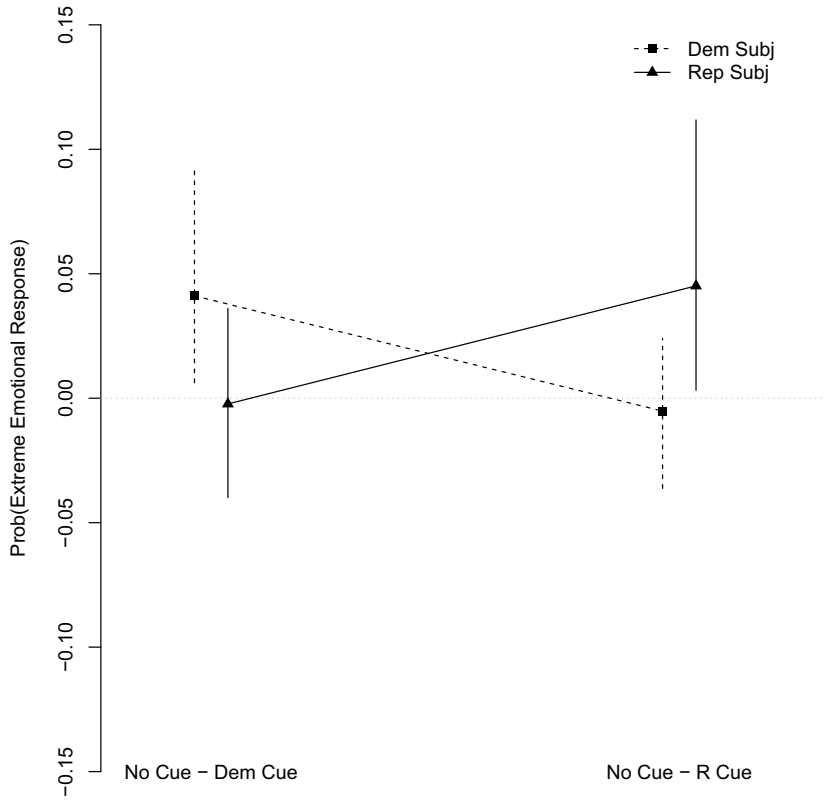

Figure 5.5: Partisan Cues and Emotional Reactions. The plot shows the predicted difference in the probability of giving an extreme emotional reaction when receiving a partisan cue. The solid line shows the predicted effect for Republicans and the Dashed lines show the predicted effects for Democrats. The vertical bars give the 95% confidence intervals.

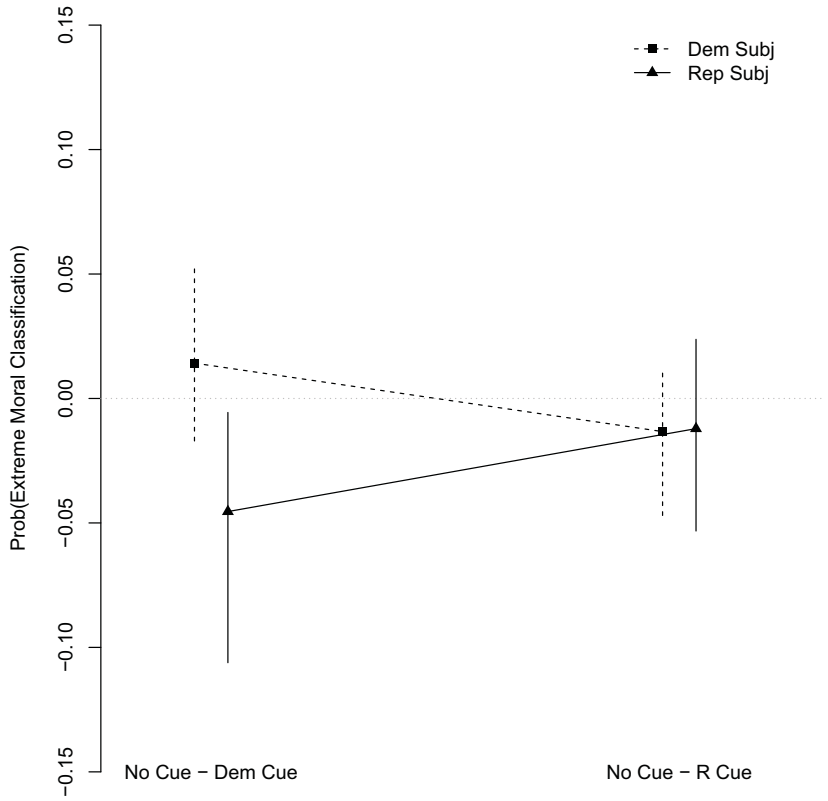

Figure 5.6: Partisan Cues and Moral Evaluations. The plot shows the predicted difference in the probability of giving an extreme moral evaluation when receiving a partisan cue. The solid line shows the predicted effect for Republicans and the Dashed lines show the predicted effects for Democrats. The vertical bars give the 95% confidence intervals.

Discussion and Conclusion

In this chapter, I have investigated some of the factors that lead people to have divergent moral judgments of the same stimuli. If we accept the premise that moral attitudes are important in politics – and there seems to be a great deal of evidence that this is the case (Skitka 2010; Ryan 2014; Tetlock et al. 2000) – we are naturally led to turn our attention to the causes of moral judgments. In the preceding analyses, I've argued that there are at least three things at work: 1) stable individual differences cause differential emotional reactions, 2) emotional reactions cause moral judgments, and 3) social identities systematically affect both emotional reactions and moral judgments.

I have shown that individuals are sensitive to morally congruent rhetoric. Differences in moral perception are at least partly explainable by appealing to individual differences in a small number of core moral domains. These individual differences in moral concern (operationalized here with the moral foundations) are systematically related to variation in the emotional reactions that individuals experience when exposed to political arguments.

My results were also consistent with the claim that emotional reactions cause moral judgments. One of the strongest predictors of individual moral classifications were their self-reported emotional reactions. These findings are consistent with an intuitionist understanding of moral judgment.[21]

Politics occurs in the context of social identities and relationships, and my results show that relevant social identities (in this case, partisanship) significantly affect emotional and moral reactions. Because source cues were randomly presented, I am more

[21]The results I present here are not able to settle questions about the direction of the causal arrow by any means. Without credibly manipulating emotions, I do not have evidence of causation.

confident about the causal story here.[22] Relevant information — knowing that the speaker is on your political "team" — seems to lower the guard. This leads to partisans reporting more intense emotional reactions when presented with statements from members of their political party.

In future work, it will be important to delve more into each of these claims. On the idea that emotional reactions cause moral judgments, it may be the case that personality factors such as emotional stability mediate the relationship between stimulus and moral judgment. If it is true that emotions are causal agents in moral judgment, any factors that influence the degree to which individuals experience their emotions are likely to have an effect on moral judgment (and by extension political tolerance and activism).

Another rich vein of future research lies in the ways in which politicians strategically use moral language to reach their supporters. As we understand more about the incentives facing politicians in their use of moral rhetoric, we might better understand the natural bounds placed on elite influence. If it is true that moral intuitions are stable and relatively fixed (as I contend), political strategy must work within the bounds set by mass dispositions. Moral rhetoric is not going away any time soon in politics. To the extent that we fail to understand the dynamics at play in individual minds and mass publics, our understanding of political behavior will be that much poorer.

[22]The mechanism by which this effect occurs remains less clear. There are at least two competing explanations. First, it could be that the seeing one's one party causes a flash of incidental affect that is then incorporated into downstream political reasoning. Under this account, we would expect to see the strongest positive effect among those people who have the warmest feelings about their political party. Alternatively, it could be that individuals are able to use the party label as a heuristic to infer more about relatively vague statements. If this alternative account were true, we might expect to see an interaction between the content of the statement and the partisanship of the speaker.