Liberals and Conservatives Rely on Very Similar Sets of Foundations When Comparing Moral Violations *

Applications of Moral Foundations Theory (MFT) in political science have revealed differences in the degree to which liberals and conservatives explicitly endorse five core moral foundations of care, fairness, authority, loyalty and sanctity. We argue that differences between liberals and conservatives in their explicit ratings of abstract and generalized moral principles do not imply that citizens with different political orientations have fundamentally different moral intuitions. We introduce a new approach for measuring the implicit importance of the 5 moral foundations by asking UK and US survey respondents to compare pairs of vignettes describing violations relevant to each foundation. We analyse responses to these comparisons using a hierarchical Bradley-Terry model which allows us to evaluate the relative importance of each foundation to individuals with different political perspectives. Our results suggest that, despite prominent claims to the contrary, voters on the left and the right of politics share broadly similar moral intuitions.

Word count: 11998

A core concern in contemporary democratic politics is whether adherents to different political ideologies might have fundamentally incompatible moral outlooks. Moral Foundations Theory (MFT) ([Haidt and Graham](2007); [Haidt](2012)) – which aims to document and explain variation in the moral perspectives of different political and social groups – suggests that differences in such intuitions may make reasoned political debate challenging. Viewing morality as residing in the intuitive reflexes that individuals give in response to moral stimuli, MFT suggests that there are five central "foundations" that inform people's moral outlooks: care, fairness, loyalty, authority, and sanctity. While these foundations are considered the "irreducible basic elements" ([Graham et al.](2013), 56) of human morality, MFT suggests that the moral weight assigned to each of these foundations by any given individual will be a function of experience, upbringing, and culture. As a consequence, MFT suggests that there

---

will be predictable differences in the moral values of those who occupy different parts of the political spectrum.

Empirical research suggests that conservatives and liberals do, in fact, differentially endorse the five moral foundations, at least when asked to reflect explicitly on their own moralities. The main evidence on which such conclusions rest comes from public opinion surveys, most of which employ the Moral Foundations Questionnaire (MFQ) (Graham et al., 2011) to measure moral attitudes. Although this literature has provided important insights into the relationship between political ideologies and moral endorsement, in this paper we argue that a number of methodological features – both of the MFQ and of the empirical literature more broadly – are likely to lead scholars to overstate differences in the moral intuitions of liberals and conservatives. We make three main arguments.

First, existing work typically relies on respondents' explicit ratings of abstract and generalized moral principles, rather than trying to assess the intuitions that implicitly structure their moral evaluations. Consistent with arguments of other scholars (e.g. Clifford et al., 2015), we argue that this approach to measuring moral priorities is in tension, however, with a core theoretical assumption of MFT. Second, the surveys used to assess the political predictions of MFT are not well-suited for drawing conclusions about the *relative* importance of each foundation for moral evaluation. Third, even when measurement approaches in this literature aim to measure implicit moral attitudes, they tend to do so by prompting respondents to provide judgments on a very small number of specific moral violations. In combination, we argue that these methodological features of existing approaches are likely to have lead prior literature to overestimate differences in the moral intuitions of liberals and conservatives.

We address these issues by introducing a new experimental design and modelling strategy which aims to measure the relative importance of the five moral foundations to respondents of different political ideologies. In our design, we ask survey respondents from the UK and the US to compare pairs of vignettes, each of which describes a single, specific violation that is relevant to one of the foundations. Rather than asking respondents to reflect explicitly on their own moral codes, we ask

them to simply compare the vignettes, and then indicate which they think constitutes the worse moral transgression. We analyse responses to these comparisons using hierarchical Bradley-Terry models and present three main findings.

We first show that, across all respondents, estimates of foundation importance based on our measure of implicit attitudes differ in important ways from existing work. Second, we show that extreme liberals and conservatives do appear to put different weights on the foundations when making moral judgments, and the differences are in the direction predicted by moral foundations theory at least with respect to the care, authority and loyalty foundations. However, our third (and most important) finding is that the ideology-by-foundation interactions that are the central focus of MFT's political predictions explain very little of the variation in moral judgments across violations. In both the UK and the US, comparing the rankings of violation severity across all of the vignettes in our experiment for respondents with different ideologies, we find large, positive correlations, even between respondents from opposite extremes of the political spectrum. We find much lower correlations applying a similar experimental design to the MFQ items, showing that ideological divisions that appear when respondents are asked to self-theorise about their morality are not present when they make intuitive moral judgments about concrete moral violations.

Our design provides a test of the core political predictions of Moral Foundations Theory. While several papers criticize MFT on the basis of theoretical objections to the theory's core concepts (Suhler and Churchland, 2011), the ambiguity between the descriptive and prescriptive components of the theory (Jost, 2012), the causal process assumed by MFT (Hatemi, Crabtree and Smith, 2019), or empirical inconsistencies with the theory's key assumptions (Smith et al., 2017; Walter and Redlawsk, 2021), we take the central ideas of MFT as a given, and ask whether political ideology is predictive of the comparisons individuals make between violations of the five foundations. Our design provides a sharper test than previous studies of the core political ideas of MFT, as our experiment is designed explicitly to solicit the types of fast, automatic, intuitive judgments that are central to the theory. By rooting our study in these types of moral judgment, we show that the political differences predicted

by the theory have been overstated in the existing literature, and that ideologically distant voters make strikingly similar intuitive moral decisions.

The basic descriptive claim that liberals and conservatives differ in terms of their moral intuitions is often used as an empirical touchstone for a wide range of studies that rely upon MFT as a theoretical approach. This claim features in work on, *inter alia*, political rhetoric (e.g. Jung, 2020; Clifford and Jerit, 2013), political persuasion (e.g. Kalla, Levine and Broockman, 2022), affective polarization (e.g. Garrett and Bankert, 2020), voting behaviour (e.g. Franks and Scherr, 2015) and public attitudes on foreign policy (e.g. Kertzer et al., 2014), "culture war" issues (e.g. Koleva et al., 2012), and immigration (e.g. Simonsen and Bonikowski, 2022). Our results – which cast doubt on how robust the liberal-conservative morality gap is to different measurement strategies – are therefore likely to be salient for many scholars working on a broad range of questions in political behaviour.

Ideological Differences in Moral Judgment

Moral Foundations Theory

How do people form moral judgments? Moral Foundations Theory (Haidt and Joseph, 2004; Haidt, Joseph et al., 2007; Haidt, 2012; Graham et al., 2013) suggests that judgments on moral issues typically arise from fast and automatic processes which are rooted in peoples' moral intuitions (Haidt, 2001). From this perspective, when encountering a situation that requires moral evaluation, people come to decisions "quickly, effortlessly, and automatically" (Haidt, 2001, 1029) without being consciously aware of the criteria they use to form moral conclusions. These intuitions are therefore considered to be the key causal factors in moral judgment, while conscious and deliberate moral reasoning – in which people search for and weigh evidence in order to infer appropriate conclusions – is thought to be employed only *post hoc*, as people search for arguments that support their intuitive conclusions.[1]

Understanding the moral decisions that people make therefore requires understanding the sources

---

[1]MFT therefore views moral reasoning as no different to other forms of reasoning in that it is likely to be motivated (Ditto, Pizarro and Tannenbaum, 2009).

of their moral intuitions. MFT adopts an evolutionary account of morality, in which the central "foundations" of peoples' moral intuitions are thought to have evolved in response to a series of broad challenges: the need to protect the vulnerable (especially children); the need to form partnerships to benefit from cooperation; the need to form cohesive coalitions to compete with other groups; the need to form stable social hierarchies; and the need to avoid parasites, pathogens, and contaminants. In response to these challenges, humans are said to have evolved distinct cognitive modules that underpin the "moral matrices" of contemporary cultures. Each challenge is thought to be associated with a distinct moral foundation – care; fairness; loyalty; authority; and sanctity – which are the "irreducible basic elements" (Graham et al., 2013, 56) needed to explain and understand the moral domain. These foundations are typically further grouped into two broader categories, where care and fairness are thought of as *individualizing* foundations (because they link closely to the focus on the rights and welfare of individuals), and loyalty, authority, and sanctity are referred to as the *binding* foundations (because they emphasize virtues connected to binding individuals together into well-functioning groups).[2]

An evolutionary origin of the foundations implies that they are innate and universally shared, at least in the sense that human minds are "organised in advance of experience" (Graham et al., 2013, 61) to be receptive to concerns that are relevant to these five criteria. However, proponents of MFT also argue that "innateness" does not preclude the possibility that individual or group moralities might be responsive to environmental influences. As Graham et al. (2013, 65) argue, "the foundations are not the finished buildings [but they] constrain the kinds of buildings that can be built most easily". At the level of individuals, while the human mind might show some innate concern for all five foundations, environmental factors – like upbringing, education, and cultural traditions – will result in different people and social groups constructing moralities that weight the foundations differently.

Moral Foundation Theory combines these arguments – that morality is pluralistic, constituted of moral concerns beyond care and justice, and expressed primarily through intuitive reactions which

---

[2] Iyer et al. (2012) propose a sixth – liberty/oppression – foundation. However, the political predictions of MFT, which are our main concern, are less clearly articulated for this sixth foundation, and so we omit it from discussion here.

are shaped by experience – to provide an explanation for observable variation in expressed morality. On this basis, MFT has been used to explain moral similarities and differences across societies, changes in moral values over time, and – crucially – variation in expressed moral values across individuals from different parts of the political spectrum.

Moral Foundations and Political Ideology

One of the most prominent applications of MFT is as an explanation for moral differences between people with different political ideologies. The core political claim made by MFT's proponents is that liberals and conservatives, in the United States and elsewhere, put systematically different weight on the different foundations (Haidt and Graham, 2007; Graham, Haidt and Nosek, 2009). An extensive empirical literature, primarily based on survey evidence that connects respondents' self-reported political ideology to questions designed to measure reliance on each of the foundations in moral decision-making, appears to offer support for this hypothesis.[3] Discussed in greater detail in the section below, the broad finding of these studies is that liberals (or those on the political left) prioritize the care and fairness foundations, while conservatives (or those on the political right) have moral systems that rely to a greater degree on the loyalty, authority, and sanctity foundations (Graham, Haidt and Nosek, 2009; Graham et al., 2011; Haidt and Graham, 2007; Kertzer et al., 2014; Enke, 2020; Enke, Rodríguez-Padilla and Zimmermann, 2022).

These results have been interpreted in stark terms. For instance, Graham, Haidt and Nosek (2009, 1030) argue that MFT accounts for "substantial variation in the moral concerns of the political left and right, especially in the United States, and that it illuminates disagreements underlying many 'culture war' issues". Similarly, Graham, Nosek and Haidt (2012, 1) suggest that "liberal and conservative eyes seem to be tuned to different wavelengths of immorality". Likewise, Haidt and Graham (2007, 99) suggest that "Conservatives have many moral concerns that liberals simply do not recognize as

---

[3]Aside from surveys, several papers focus on describing the different sets of moral appeals used in texts produced by individuals and groups with different ideological dispositions, such as religious leaders (Graham, Haidt and Nosek, 2009), voters (Feinberg and Willer, 2015), students (Rempala, Okdie and Garvey, 2016), and newspapers (Clifford and Jerit, 2013).

moral concerns." Together, the impression generated by this literature is that there is a fundamental incompatibility between the moral outlooks of liberals and conservatives.

MFT clearly envisages the relationship between morality and political ideology to be causal, with moral intuitions shaping political stances (Haidt, 2012; Koleva et al., 2012; Kertzer et al., 2014; Franks and Scherr, 2015; Enke, 2020; Enke, Rodríguez-Padilla and Zimmermann, 2022). However, recent work has questioned this account by showing that changes in political ideology predict changing moral attitudes, rather than the reverse (Hatemi, Crabtree and Smith, 2019). Similarly, in contrast to the stable, dispositional traits required to be convincing determinants of political ideology, moral attitudes are subject to substantial individual-level variability over time (Smith et al., 2017). Survey experimental evidence also suggests that endorsement of the different moral foundations is sensitive to political and ideological framing effects (Ciuk, 2018), and the morality gap between liberals and conservatives is sensitive to other attitudes of individuals, such as how closely aligned a person is with their social group (Talaifar and Swann Jr, 2019) and how politically sophisticated they are (Milesi, 2016).

While these papers complicate the causal story told by MFT, they do not dispute the idea that, descriptively, liberals and conservatives endorse different sets of moral principles. As Hatemi, Crabtree and Smith (2019) suggest, "our findings provide reasons to reconsider MFT as a causal explanation of political ideology [but] do nothing to diminish the importance of the relationship between these two concepts." More generally, throughout the empirical literature on MFT, disagreement focuses on how to *interpret* the correlation between moral and political attitudes, not on whether such a correlation exists or how large it is. Regardless of the overarching causal story, interpretations of the available descriptive evidence are relatively uniform: there are substantial differences in the moral outlooks of liberals and conservatives, both in the US and elsewhere.

Measuring Foundation Importance

The strength of the descriptive association between moral and political attitudes is likely, however, to be related to the instruments used for measuring individuals' reliance on the five moral foundations. Existing empirical research relies heavily on The Moral Foundations Questionnaire (MFQ) (Graham, Haidt and Nosek, 2009; Graham et al., 2011), a survey instrument designed to assess individual-level endorsement of the moral foundations, which is composed of two question batteries. The first battery asks participants to rate how relevant various concerns are to them when making moral judgments.[4] The "moral relevance" items used in this battery are typically abstract and generalized statements such as "whether or not someone was harmed" (care), or "whether or not someone did something disgusting" (sanctity). The second battery aims to assess levels of agreement with more specific "moral judgments" by asking participants to rate (from strongly disagree to strongly agree) their agreement with statements such as "respect for authority is something all children need to learn" (authority) or "when the government makes laws, the number one principle should be ensuring that everyone is treated fairly" (fairness). These two batteries have been used extensively throughout existing research which measures political differences in moral endorsement (e.g. Graham, Haidt and Nosek, 2009; Graham et al., 2011; Haidt, 2012; Koleva et al., 2012; Kertzer et al., 2014; Franks and Scherr, 2015; Enke, 2020; Enke, Rodríguez-Padilla and Zimmermann, 2022).

Despite the ubiquity of this survey instrument, the MFQ is subject to a number of shortcomings. First, MFT assumes that moral *intuitions* – fast, effortless reactions to moral stimuli – are central to moral decision-making. Crucially, the cognitive processes that lead to intuitive reactions are thought to be inaccessible to respondents. Moral intuitions are marked by the sudden appearance of moral conclusions in the mind, "without any conscious awareness" of having gone through the process of forming an opinion, nor any recognition of the factors that lead to a particular conclusion being reached (Haidt, 2001, 1029). However, while MFT assumes the primacy of intuition, the vast majority

---

[4]"When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking?"

of the survey questions used to assess morality differences between liberals and conservatives are based on questions that solicit slow, self-reflective, "System-2" style, responses. The MFQ prompts respondents to self-assess their own motivations for their moral choices; motivations which are – by MFT's own assumptions – inaccessible to them. The gap between theoretical assumptions and empirical methods in MFT research represents, for some, "a substantial impediment to testing and developing theories of morality." (Clifford et al., 2015, 1179)

Prompts asking respondents to explicitly endorse moral values may lead to a different distribution of responses than items that elicit more automatic moral judgments. In particular, respondents who are prompted to self-theorise about their own moralities may be more likely to engage in motivated moral reasoning (Ditto, Pizarro and Tannenbaum, 2009), stressing the moral principles that are most supportive of the positions they take on particular moral issues. This creates a pathway for spurious political differences to arise in MFQ responses as respondents give the responses that make sense of their politics. However, political differences in the ways in which people *justify* their moral choices are distinct from differences in the ways that people intuitively *evaluate* specific moral scenarios. This argument is shared by Graham, Haidt and Nosek (2009, 1041), who suggest that "studies using implicit measurement methods will be essential for understanding the ways in which liberals and conservatives make moral judgments."

In addition, the items included in the MFQ typically describe abstract moral principles, which respondents may interpret differently. When asked to rate the moral relevance of "Whether or not someone showed a lack of respect for authority", for instance, people might imagine very different scenarios, which may differ in moral importance. One respondent might imagine a situation in which a child is rude to a parent, while another might imagine a solfier refusing to follow the instructions of their commanding officer. In general, because the MFQ fails to prompt respondents to consider specific moral violations, it risks respondents imputing scenarios that they associate with general categories of moral wrong, and the scenarios that respondents impute might differ systematically across respondents with different ideologies.

Second, the survey prompts used in this literature do not easily facilitate comparisons of the relative importance of the different foundations to moral decision-making. In particular, the MFQ contains sets of items which respondents are asked to rate one at a time, against an abstract scale. In contrast, the pairwise comparison approach we adopt facilitates the intuitive responses that are most theoretically relevant to MFT, because they do not require comparison of each item against an abstract scale that must be first theorised and then mentally carried across items and applied to each. By avoiding this cognitive burden, pairwise comparisons also facilitate consistent responses for an individual across prompts. Several authors have argued that future empirical work on MFT should consider adopting comparison-based, rather than rating-based, questions in order to encourage respondents to directly consider trade-offs between different foundations (Ciuk, 2018; Jost, 2012).

Third, even when researchers have used specific moral violations (rather than abstract moral principles) as the basis of inference for political differences in expressed morality, they have tended to rely on a small number of examples which are supposed to be representative of each foundation. For example, three prominent examples of sanctity violations – one involving incest, one involving eating a dead dog, and one involving having sex with a dead chicken – have been used extensively in a number of studies relating to moral judgment (Wheatley and Haidt, 2005; Schnall et al., 2008; Eskine, Kacinik and Prinz, 2011; Feinberg et al., 2012). More generally, as Clifford et al. (2015, 1180) argue, while researchers have used a variety of vignettes in past work on MFT, none of these vignettes alone "cover the full breadth of the moral domain." Reliance on a small number of issues might lead to overestimates of political differences in moral judgment if the issues selected are marked by unusually large levels of partisan disagreement. As Frimer et al. (2013, 1053) note, the examples used in previous research may be "unrepresentative of the full spectrum of moral judgments that people make."

Experimental designs in which researchers study the effects of latent concepts (such as the moral content of a given scenario) using a small number of specific implementations are subject to potential confounding concerns. The scenarios that researchers construct may differ from each other in multiple ways, not only in terms of the latent treatment concept they are intended to capture (Grimmer

and Fong, 2021). Differences in moral foundation endorsement may not result from differences in moral content, but rather from the fact that the researcher-generated MFT scenarios used to typify each foundation are confounded by differing levels of "weirdness and severity" (Gray and Keeney, 2015, 859). In addition, studies that use single-implementations of latent treatment concepts tend to have low levels of external validity, as the treatment effects of one implementation of a given latent treatment may differ in both sign and magnitude from the effects of another implementation of the same concept (Blumenau and Lauderdale, 2022; Hewitt and Tappin, 2022).

Finally, most of the studies that document morality differences between liberals and conservatives are based on self-selected, convenience samples which are unlikely to be representative in terms of political ideology or other covariates (Graham, Haidt and Nosek, 2009; Graham et al., 2011). Although sample selection is only consequential for conclusions about the political dimension of moral endorsement when there are interaction effects between participation decisions, political orientations and expressed morality, this is nevertheless an aspect of existing designs where there is room for improvement.

Together, these features of existing survey measurement approaches suggest that moral foundation-based differences between political liberals and conservatives may have be overstated in the previous literature on MFT. In the next section, we propose a new experimental design and modelling strategy that aims to elicit the types of intuitive responses that are central to MFT; emphasizes comparative evaluations of the different dimensions of morality; deploys a large number of specific, concrete examples of violations associated with each foundation; and constructs estimates for nationally representative samples in the UK and the US.

Experimental Design

In this section we describe the design of two survey experiments. The first experiment asked respondents to make choices between pairs of moral foundation vignettes (MFV) which describe specific violations of particular moral foundations. This experiment was fielded by YouGov to samples of US

and UK respondents in February 2022. To sharpen the contrast between the results from this survey and those from the existing literature, we conducted a second pairwise-comparison experiment (also fielded by YouGov) in April 2023, in which a sample of US respondents compared pairs of items from the MFQ itself.[5] While the treatment texts we use vary across these experiments, the experimental design and modelling strategy we use for analysis are common to both.

Moral Foundation Vignettes (MFV)

The design of our first experiment is based around 74 short vignettes, each of which describes a behavior that violates one specific moral foundation. The vignettes we use are drawn from Clifford et al. (2015), who develop the texts with the goal of providing standardized stimulus sets which map directly to each foundation. Each vignette describes a situation "that could plausibly occur in everyday life" (Clifford et al., 2015, 1181) and the vignettes are written to minimize variability in both text length and reading difficulty. In order to maintain the distinction between moral and political intuitions, the violations avoid any "overtly political content and reference to particular social groups" (Clifford et al., 2015, 1181). Clifford et al. (2015) show that survey respondents associate these vignettes with the foundations to which they are intended to apply, and that respondents' perceptions of the moral wrongness of these vignettes correlate broadly with their answers to the MFQ.

We lightly edited these vignettes for use in our context. In particular, as we fielded them to respondents in both the UK and the US, we changed the wording of some vignettes to make them consistent with the idioms and political contexts of each country.[6] We also use two versions of each vignette: one in which the person committing the moral violation is a man, one where it is a woman.[7] Finally, we also removed 5 of the Clifford et al. (2015) vignettes entirely from our sample because they did not translate into realistic scenarios in both countries.[8] The sample of violations is not balanced

---

[5]We fielded the second experiment in the US only because the most prominent work on the relationship between morality and political ideology focuses on US samples and, as discussed below, we observe very little cross-national heterogeneity in the relationship between ideology and moral judgment in response to the MFV items.

[6]For example, "A woman leaving her dog outside in the rain after it dug in the [**trash**/**rubbish**]".

[7]For example, "A [**woman**/**man**] leaving [**her**/**his**] dog outside in the rain after it dug in the trash".

[8]We also excluded one vignette about having intimate relations with a recently deceased loved one because we thought

across the 5 foundations: we have 26 care foundation violations, 12 fairness violations, 14 authority violations, 13 loyalty violations, and 9 sanctity violations. We present all vignettes in table 1 in the appendix.

The key virtue of the treatments we include in our MFV experiment is that each vignette describes a specific action that constitutes a violation of a specific foundation. Accordingly, rather than asking respondents to reflect on the importance of each foundation to their own moral reasoning (as in the MFQ), we instead try to infer the degree to which respondents rely on the different foundations by examining their judgments of these scenarios. This approach is consistent with the idea that respondents' theories of their own moralities may differ from the ways in which they actually make intuitive moral judgments, and it is the latter that are central to the political predictions of MFT.

An additional benefit of our approach is that we use a wide range of vignettes to operationalise violations of each of the five foundations. Using multiple violations relevant to each foundation reduces the risk that our inferences will be skewed by confounding factors present in any specific treatment implementation, a common problem in survey-experimental designs employing text-based treatments (Grimmer and Fong, 2021; Blumenau and Lauderdale, 2022). Of particular concern is that any given implementation of a foundation-specific violation may be subject to especially pronounced political differences. For instance, if we used a single-implementation design and our example of a sanctity violation related to the issue of abortion, we might expect very large political differences that are not necessarily representative of typical differences in the moral intuitions of liberals and conservatives relating to sanctity concerns. By using a large number of violations of each foundation, we reduce the risk that the differences we estimate will be attributable to the idiosyncrasies of any particular treatment implementation.

Finally, it is important to note that the set of violations we use are not in any sense a representative sample from a well-defined population of violations of each type. In fact, it is not clear that it would be possible even to characterize such a population. However, the large number of violations (both

---

it pushed the limits of what constitutes an acceptable survey question.

overall and of each foundation) and the variation in the substance of each of the scenarios is consistent with calls to use "a broader set of cases to explore judgments of right and wrong" (Clifford et al., 2015, 1179) in the process of evaluating MFT. We discuss the differences and similarities between the results in Clifford et al. (2015) and the results from our experiment in detail below.

Moral Foundation Questionnaire (MFQ)

The design of our second experiment is based on the 30-item version of the MFQ (Graham et al., 2011) which has been used extensively in the empirical literature on MFT. The items in the MFQ are of two types. First, there are 15 "moral relevance" items that seek respondents' views on a set of abstract statements about the relevance of particular moral values. Second, there are 15 "moral judgment" items which constitute a mixture of normative declarations, virtue endorsements, and opinions about the principles of government policy (Graham et al., 2011, 371), with which respondents are asked to indicate how strongly they agree or disagree. We list all the MFQ items in appendix A2.

The core advantage of using the MFQ items in this second experiment is that it allows us to directly compare the ways in which respondents make judgments of concrete moral violations, and the ways in which they make judgments between more abstract moral statements (which form the basis of the existing empirical literature), while holding the paired-comparison survey design and modelling approach that we use constant.

Randomization, Prompts and Sample

In both experiments, we present randomly-selected pairs of items – either from the MFV or the MFQ – to survey respondents and ask them to select between them. For the first experiment, we ask respondents to select which of a given pair of violations is "more wrong".[9] In the example given in figure 1, the respondent sees one vignette about a man lying about the number of days he has taken off from work (a violation of the "fairness" foundation), and one vignette about a man who punished his

_____

[9]We provide information on our opt-out pre-screen and ethics approvals in appendix A6.

Which of these do you think is more wrong?

A man lying about the number of paid days off he has taken from work.	A man leaving his dog outside in the rain after it dug in the rubbish.

They are about the same

Figure 1: Experimental prompt, Moral Foundation Vignette (MFV) items.

Please read the following two sentences and indicate which you agree with more.

Men and women each have different roles to play in society.	**People should be loyal to their family members, even when they have done something wrong.**

I agree or disagree with these equally

When you decide whether something is right or wrong, which of the following is more relevant to your thinking?

Whether or not someone showed a lack of loyalty	**Whether or not someone did something disgusting**

Both are equally relevant or irrelevant

Figure 2: Experimental prompts, Moral Foundation Questionnaire (MFQ) items.

dog for bad behavior by leaving it out in the rain (a violation of the "care" foundation). Respondents were able to click on which of the two vignettes they thought was worse, or alternatively could select "They are about the same".

In the second experiment, half the respondents were presented with randomly-selected pairs of items from the "judgment" item battery and asked to indicate which item they agreed with more, and the other half of respondents were were presented with pairs of items from the "relevance" item battery and were asked to indicate which of the items was "more relevant to your moral thinking". In both cases, respondents could also select an intermediate option that favoured neither treatment text. Examples of both forms of prompt are given in figure 2.

For each comparison, we first sampled two foundations and then, conditional on the foundation drawn we sampled two of the vignettes/items, without replacement, from the full set of 74 MFV violations or, in the MFQ experiment, from the 15 items relevant to a respondent's treatment arm

(i.e. relevance items or judgment items). This sampling strategy means that we have equal numbers of observations of each *foundation* in expectation, but unequal numbers of observations for each *violation* (because we have more violations for some foundations than for others). For each of the MFV vignettes, we also randomly sampled whether the person committing the relevant violation was a man or a woman. In both experiments, each respondent answered 6 pairwise comparisons. For experiment one, we collected data from 1598 respondents in the UK and 2375 respondents in the US, giving us a total of 9472 and 14063 observations from the UK and US respectively.[10] In experiment two, we collected data from 2321 respondents in the US, giving us a total of 13926 observations.

Measuring Moral Intuitions

Model definition

Each of our experiments result in an ordered response variable with three categories. We present the models below with reference to our first (MFV) experiment; we use an identical model to investigate the response distributions of our second (MFQ) experiment. For experiment one, we have:

$$Y_i \in \begin{cases} 1 = \text{Violation 2 is worse} \\ 2 = \text{About the same} \\ 3 = \text{Violation 1 is worse} \end{cases} \tag{1}$$

To model this outcome, we adopt a variation on the Bradley-Terry model for paired comparisons (Bradley and Terry, 1952; Rao and Kupper, 1967) where we model the log-odds that violation j is worse than violation j' in a pairwise comparison:

$$log\left[\frac{P(Y_i \leq k)}{P(Y_i > k)}\right] = \theta_k + \alpha_{j(i)} - \alpha_{j'(i)} \tag{2}$$

---

[10] For the first experiment, we have 116 missing outcome responses from our UK sample and 187 from our US sample. These are cases where survey respondents failed to complete a given comparison task. We have no missing outcome responses for the second experiment.

where θ_k is the cutpoint for response category k. We can interpret the α_j as the "severity" of violation j (or, equivalently, as the "relevance" or "importance" of item j in the MFQ experiment). This parameter is increasing in the frequency with which respondents choose violation j as the "worse" moral violation in paired-comparison with other violations.

In this section, our primary goal is to understand how the severity of these violations varies according to which of the 5 moral foundations they violate. That is, we are not primarily interested in the relative severity of the 74 individual moral scenarios, but rather in how the distribution of severity differs for violations of different types. We have a moderate number of observations for each of the violations that we include in our experiment: on average, each vignette appears in 256 comparisons in the UK data and 380 comparisons in the US data. As a consequence, our design is only likely to be well-powered to detect reasonably large differences in average moral evaluations of the individual *violations.* However, we have far more information about the average moral evaluations of the different *foundations* and about the levels of variation across vignettes, which are our main targets of interest. We therefore use a hierarchical approach to estimating the average and distribution of effects of each of the five foundations by specifying a second-level model for the α_j parameters.

Where $f(j)$ is the moral foundation violated in vignette j, we model the severity of each violation as the sum of a foundation effect, $\mu_{f(j)}$, plus a violation-specific random-effect, ν_j:

$$\alpha_j = \mu_{f(j)} + \nu_j \tag{3}$$

The model described by equation 3 implies that the severity of a given violation equals the average severity of violations of a given foundation type plus a deviation that is attributable to that specific violation. One advantage of our modelling approach is that it allows us to quantify the distribution of violation severity for each foundation.

The model is completed by prior distributions for the ν_j and $\mu_{f(j)}$ parameters, which we assume are drawn from normal distributions with mean 0 and standard deviation σ_ν and σ_μ, respectively. We estimate a single scale parameter for the distribution of violation-effects, regardless of the foundation

to which they apply. We estimate the model using Hamiltonian Monte Carlo as implemented in Stan (Carpenter et al., 2017). The results presented below are based on 4 parallel simulation chains of 1000 iterations which follow from 500 warm-up iterations.

Finally, as we argued above, existing survey evidence in support of MFT is largely based on convenience samples that are unlikely to be representative. We maximize the representativeness of our estimates from both experiments by incorporating demographic survey weights, provided by YouGov, via a quasi-likelihood approach. The estimates are substantively identical to estimates produced without using the weights.

Results

We present the main results from this baseline model in figure 3. The figure shows the estimated average severity of violations of each of the five moral foundations (μ_f) from experiment one in both the UK (left panel) and the US (centre panel), as well as the average relevance of the five moral foundations from experiment two (right panel). It also depicts the estimated severity/relevance of each of the 74 individual MFV violations and 30 MFQ items ($\mu_{f(j)} + \nu_j$) that we include in the two experiments (transparent points and intervals). Numerical estimates are provided in appendix A7.

The figure reveals two main findings. First, we recover systematic differences in the average severity of the tested violations across the different foundations, and the ordering of the importance of the foundations differs between the MFV and MFQ versions of the experiment. In particular, the figure clearly demonstrates that the violations of the care, fairness, and sanctity foundations that we tested are, on average, considered to be morally worse by our respondents – in both the UK and the US – than violations of either the authority or loyalty foundations. By contrast, the MFQ experiment suggests that care and fairness are more relevant to respondents' moral thinking than all three of the other foundations, including sanctity.

This ordering of the moral wrongness of violations of different types is consistent with Clifford et al. (2015, 1188), but contrasts with the generally accepted ordering of moral foundation importance
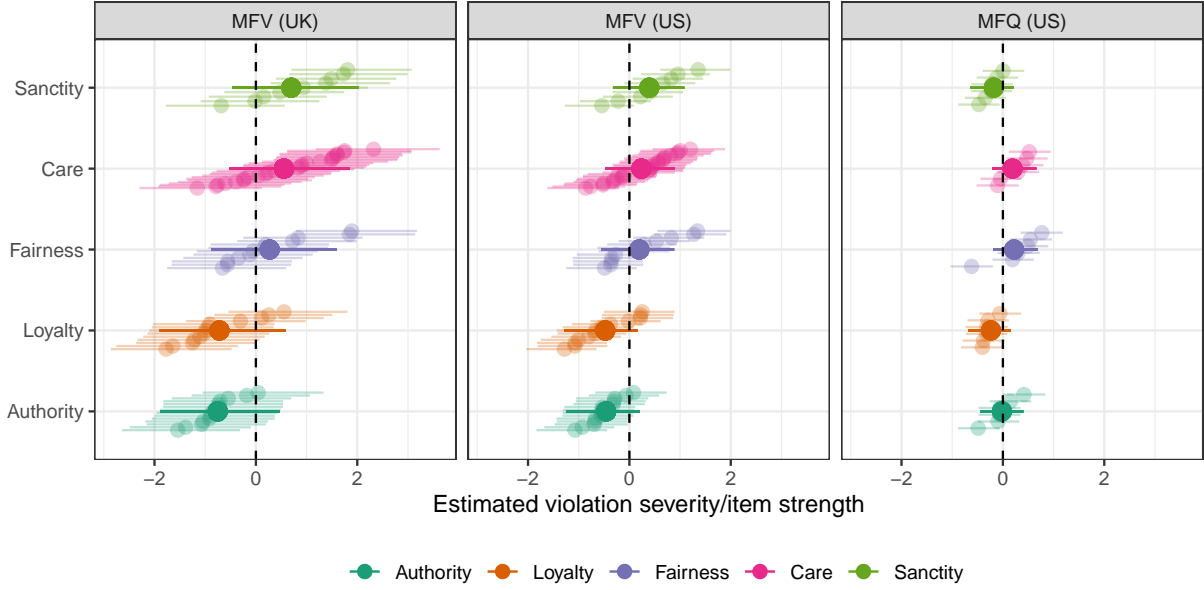
Figure 3: Estimates of $\mu_{f(j)}$ and α_j from equations 2 and 3. Left and centre panels give estimates from the UK and US versions of the MFV experiment, respectively. The right panel gives estimates from the MFQ version of the experiment.

in the literature. For instance, Graham, Haidt and Nosek (2009, 1032) find that, averaging over individuals of different political positions, the moral relevance of the "individualizing" foundations of care and fairness is significantly higher than that of the "binding" foundations of loyalty and authority. However, our finding that respondents also rate violations of the sanctity foundation, on average, as worse than either loyalty or authority violations, and roughly equally as bad as care and fairness violations contrasts with findings presented in Graham, Haidt and Nosek (2009) and Haidt (2012), where – even among the most conservative respondents – sanctity considerations are considered less important to moral decision-making than either harm or fairness concerns.

One interpretation of this difference is that it reflects the different design choices between our two experiments (and between those in our violations experiment and the existing literarture). When survey respondents are asked to reflect in the abstract on the considerations that are most important to them in their moral decisions, they tend to think that sanctity concerns are not as important as other moral criteria. But if you ask respondents to compare concrete examples of human action, those that describe degrading (but harmless) situations are often selected as the worst violations of

acceptable moral behavior.

An alternative interpretation is that the sanctity violations included in Clifford et al. (2015) are more severe among the set of all possible sanctity violations than was the case for the violations we presented among those possible for the other foundations. A parallel concern exists for the MFQ as well, as it is not at all clear what severity of violation respondents are imagining when asked to self-assess how responsive they are to a type of violation in the abstract, or indeed how sensitive the measures are to the wording of the abstract violation items. As noted above, and returned to below, it is difficult to define a population of violations, concerns or principles in a way that would allow one to ensure representativeness and thus make reliable claims about which foundations are more important overall.

This observation leads to our second main finding of this initial model: we observe significant heterogeneity in violation severity both across and within foundation types. First, the overall range of the MFV violation severity estimates dwarfs that of the MFQ item strength estimates, implying that respondents are significantly more decisive when it comes to judging which violations are worse than they are about which items of the MFQ are more relevant to their moral thinking.[11] Second, in the MFV experiment, there is a large degree of overlap in the severity of violations of different types. Only a limited fraction of the variation in respondents' evaluations of severity can be explained by the foundations to which the violations apply. The foundation effects ($\mu_{f(j)}$) explain 37% of the variation in violation severity (α_j) for UK respondents and 31% of the variation for US respondents. The same is true for experiment two, where the foundation effects explain 31% of the variation in MFQ item relevance.[12] The foundation to which the MFV vignettes and MFQ items are associated is predictive of respondents' judgments, but approximately 2/3rds of the variation across these remains unexplained by foundation.

---

[11]This is also evident from the pattern of intermediate responses. In the MFV experiment, respondents give the "They are about the same" response in 26% of comparisons in the UK, and 32% of comparisons in the US. In the MFQ experiment, respondents give intermediate responses in 45% of comparisons.

[12]We follow Gelman and Pardoe (2006) to calculate the R^2 at the second level of the model: $R^2 = 1 - \frac{E(V_{j=1}^J \nu_j)}{E(V_{j=1}^J \alpha_j)}$, where E is the posterior mean and V is the variance.

In appendix A3, we present the violations considered most and least severe for each foundation, in each country. The violations that feature in the US and UK lists in the tables are very similar, suggesting a high degree of correlation in moral evaluations across the two countries in our sample.

Measuring Moral Intuitions by Ideology

Model definition

The model above allows us to measure the extent to which the different moral foundations predict respondents' judgments between pairs of moral violations or pairs of abstract moral statements. However, the central political claim made by MFT is that the relevance of each of the foundations to moral decisions will depend on the ideological position of a given individual, and this model does not yet allow us to describe how these foundation-level effects vary by respondent ideology.

We therefore modify the model in equations 2 and 3 to allow the violation-level parameters, α_j, to vary according to the self-reported ideological position of the respondent. As before, we describe this model in relation to the MFV experiment but implement an identical model for the MFQ experiment.

We follow Graham, Haidt and Nosek (2009) and ask all respondents to place themselves on a seven-point ideological scale before they complete the violation comparison task.[13] We include this variable in a model of the following form:

$$log\left[\frac{P(Y_i \leq k)}{P(Y_i > k)}\right] = \theta_k + \alpha_{j(i),p(i)} - \alpha_{j'(i),p(i)} \tag{4}$$

where $\alpha_{j,p}$ is the severity of violation j for ideology-group p. We then model these parameters with an adapted second-level model in which we allow the foundation effects, $\mu_{f(j)}$, to also vary by respondent

---

[13]Respondents in the US select from {*Strongly Liberal, Liberal, Slightly Liberal, Moderate, Slightly Conservative, Conservative, Strongly Conservative*}. UK respondents select from {*Strongly Left, Left, Slightly Left, Moderate, Slightly Right, Right, Strongly Right*}. 316 UK individuals selected "Don't know" for this question and 200 individuals in our US sample. To avoid dropping these respondents, we recode them as "moderate" on our ideology variables. None of our results are sensitive to this choice.

ideology:

$$\alpha_{j(i),p(i)} = \mu_{f(j)} + \gamma_{f(j),p(i)} + \nu_j \tag{5}$$

In this specification, $\gamma_{f,p}$ is a matrix of coefficients which describe how the main effects of foundation severity (μ_f) vary as a function of the ideology of the respondent making the comparison between violations. That is, $\gamma_{f,p}$ collects the set of foundation-by-ideology interaction effects that are central to the political claims made by MFT. To identify the model, we set one foundation – fairness – as the baseline category ($\mu_{fairness}$ is constrained to be zero). As for the model in the previous section, we assume a normal prior for ν_j, with mean zero and standard deviation σ_ν. We assume improper uniform priors for μ_f.

For the interaction effects for each foundation, we use a first-order random-walk prior, such that the effect for a given ideology group on a given foundation is drawn from a distribution with mean equal to the effect for the adjacent ideological group and standard deviation σ_γ:

$$\gamma_{f,p} \sim N(\gamma_{f,p-1}, \sigma_\gamma) \tag{6}$$

We estimate all the ideology-level interaction effects, $\gamma_{f,p}$, relative to the "moderate" group of respondents, meaning that $\gamma_{f,4} = 0$ for all foundations. The random-walk prior encourages smooth coefficient changes between adjacent ideological groups, unless the evidence from the data is sufficiently strong to indicate otherwise.

Results

We present the estimates of the foundation-level effects for each level of ideology ($\mu_f + \gamma_{f,p}$) from this model in figure 4.[14] The figure reveals important differences between our two experiments in relation to the association between ideology and moral judgment. The results from the MFQ experiment in

---

[14]Numerical values for these estimates are given in appendix A7.

the bottom row broadly replicate classic findings from the literature on moral foundations. When comparing abstract moral ideals, liberals care primarily about the fairness and care foundations, while conservatives appear to have a "broader moral matrix" Haidt (2012, 357), placing similar weight on all five foundations. The figure also reveals steep ideological gradients for all five foundations in the MFQ experiment, with concern for authority, loyalty, and sanctity all sharply increasing when moving from liberal to conservative respondents, and concern for fairness and care decreasing among right-wing respondents.

By contrast, the top and middle rows of the figure demonstrate that the ideological associations are much weaker when estimated on the basis of responses to the moral violation vignettes. Compared to the estimates based on the MFQ, the ideological associations are noticeably shallower across the sanctity, care, fairness and authority foundations, while the relationship between ideology and loyalty is broadly similar in the two experiments. For two of the foundations – fairness and sanctity – we find no relationship at all between ideology and moral judgments. Although there is some indication that the most extreme liberal respondents in the US put greater weight on fairness violations than other respondents, there is no difference between the other 6 ideological categories on this dimension, and ideology does not affect perceptions of fairness violations in the UK either. Similarly, with respect to sanctity, although thought to be a key dimension on which liberals and conservatives differ (Graham, Haidt and Nosek, 2009; Haidt and Graham, 2007), we find that no group places significantly more weight than any other group on violations of this type when making moral judgments.

One potential caveat relating to our null result on the fairness foundation is that the fairness vignettes developed in Clifford et al. (2015) are arguably more closely linked to the notion of fairness as "proportionality" – the desire to see people rewarded or punished in proportion to the moral quality of their deeds – rather than fairness as "equality" – a form of social reciprocity marked by equal treatment, equal opportunity and equal shares (Clifford et al., 2015, 1193). The original conceptualization of the fairness foundation combined elements of both proportionality and equality, but later revisions have suggested that this foundation should be primarily about proportionality (Haidt, 2012, 209). So

conceptualized, the prediction of MFT is that "[e]veryone – left, right, and center – cares about pro-portionality…But conservatives care more, and they rely on the Fairness foundation more heavily – once fairness is restricted to proportionality." (Haidt, 2012, 213) This prediction contrasts with earlier work in MFT which suggested that the fairness foundation will be primarily associated with those on the left than on the right (Graham, Haidt and Nosek, 2009). However, even if the items developed in Clifford et al. (2015) more closely correspond to the notion of proportionality, then our results still challenge the claims of MFT, as we show that there is no systematic relationship between ideology and judgments of moral violations of this type.[15]

Moreover, when making paired comparisons of the MFV items, liberals and conservatives appear to have very similar rank orderings of the relative severity of violations of the five foundations. This is because even where we do detect ideological differences in moral judgments, these are in general much smaller than the foundation-level variation in violation severity. Those on the political right may object less to violations of the care foundation than those on the left, but the right still views the care violations we presented as substantially worse, on average, than either the loyalty or the authority violations. Likewise, even if the right put marginally more weight on loyalty considerations than those on the left, they nevertheless ranked the loyalty violations as much less severe than those involving fairness, care, or sanctity, on average. As a result, the foundation-level effects from the moral violations experiment imply that all respondents, regardless of ideology, rank the tested set of sanctity, fairness and care violations as systematically more important than the authority or loyalty violations. This again contrasts sharply with the results of the MFQ analysis, in which liberals and conservatives report very different rankings of the foundations.

An important implication of these results is therefore that estimates of political differences in moral judgment depend heavily on whether survey items aim to capture explicitly stated moral prin-ciples versus implicit moral responses to concrete situations. When asked to articulate theories of

---

[15]In addition, figure 6 below suggests that these null effects are not masking ideological heterogeneity at the level of individual violations. The left-right correlation in judgments of the individual fairness violations is very close to one, suggesting that liberals and conservatives do not differentially weight fairness violations of different types.
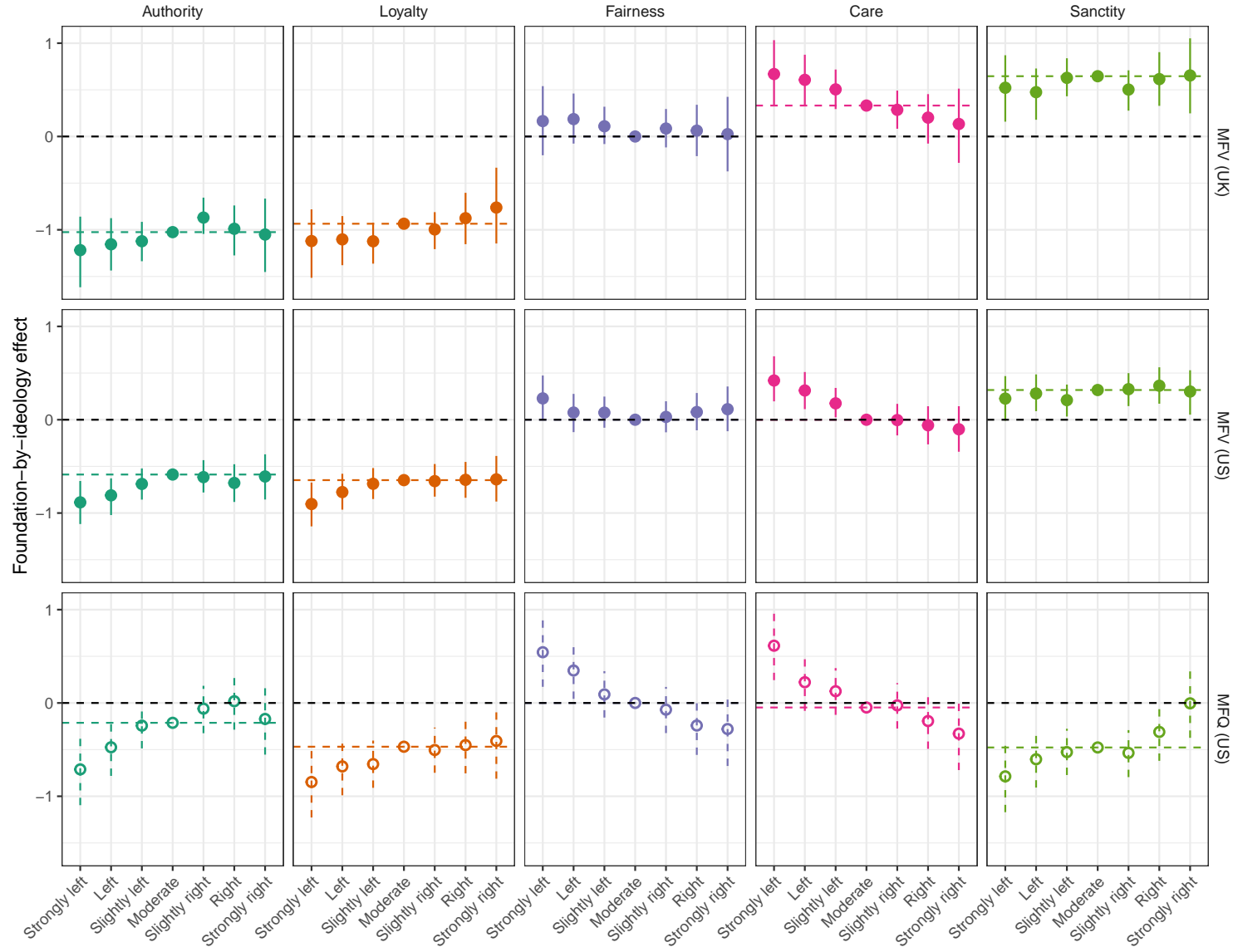
Figure 4: Estimates of $\mu_f + \gamma_{f,p}$ from equation 5. Top and Middle rows: MFV experiment, UK and US respondents. Bottom row: MFQ experiment, US respondents only.

their own morality, conservatives cite both authority and loyalty as central concerns, but when faced with specific violations of those norms they appear to view them as less important than actions that violate principles of care or fairness. Similarly, liberals might say, when asked, that they do not think sanctity concerns are relevant to their moral evaluations, but they still object when presented with scenarios in which people engage in activities that transgress ideas of temperance, chastity, and cleanliness.

Our findings share some similarities with the analysis reported in Clifford et al. (2015, 1193). Like us, Clifford et al. (2015) show that there is no correlation between ideology and judgments of fairness violations; and that the relationship between ideology and authority, and ideology and loyalty, is slightly weaker than the same relationship when measured from the MFQ. By contrast, in their analysis, Clifford et al. (2015) show a stronger relationship between ideology and the care foundation in the MFV than the MFQ (at least with respect to violations involving physical harm), where we find a weaker relationship between ideology and judgments of care violations. Finally, Clifford et al. (2015) find a strong relationship between sanctity concerns and ideology, while we find no evidence of such a relationship.

We attribute these differences largely to differences in survey design and sample composition. The analysis in Clifford et al. (2015) is based on 416 responses to a non-representative survey of 18-40 year olds in the US, in which respondents rated the moral wrongness of 132 vignettes using a 5-point scale, completed the 30-item MFQ, and answered a series of political and demographic questions. These design decisions are clearly appropriate for surveys aimed at developing and validating new items for moral judgment, but are not necessarily optimal for data collection which aims to test subsequent hypotheses about the relationship between political ideology and moral decision making. In addition to using two large and nationally representative samples, and a significantly simpler survey instrument, our analysis also differs from Clifford et al. (2015) in that we ask respondents to make comparisons between vignettes rather than providing single-vignette ratings. We therefore see our paper as a helpful extension to the work of Clifford et al. (2015), as we build directly on their work to

explore in detail the relationship between political ideology and moral judgments using the vignettes that they developed.

Measuring Agreement about Individual Violation Severity

Model definition

Given the small foundation-level differences in the MFV experiment uncovered above, we might also be interested in whether there are large political differences regarding specific violations, regardless of the foundation to which they apply. If conservatives and liberals react to the world through fundamentally different moral intuitions, then we might expect there to be little similarity in their relative assessments of specific moral violations, even if these differences do not map neatly onto the foundation-based categorization proposed by MFT.

In order to directly estimate the correlation between the relative judgments of violation severity across different ideological groups at the vignette level, we again build on the first-stage model described in equation 4. We use a "correlated severity" model where we model the $\alpha_{j,p}$ parameters by assuming that they are drawn from a multivariate normal distribution with mean zero and covariance matrix Σ:

$$\alpha_{j,p} \sim MVN(0, \Sigma) \tag{7}$$

Here, Σ has diagonal elements σ_p^2 and off-diagonal elements $\sigma_p \sigma_{p'} \rho_{p,p'}$. The correlations ρ are our primary interest, as these tell us whether the relative severity of the violations, across our entire MFV experiment, tend to be very similar for groups p and p' ($\rho_{p,p'} > 0$), whether the violations that are considered to be bad by one group are uncorrelated with those that are considered bad by the other ($\rho_{p,p'} \approx 0$), or whether the groups systematically disagree about which violations are worse from a moral perspective ($\rho_{p,p'} < 0$). We estimate this model twice, once using the seven-category version of the ideology scale for each country described above, and once using a simplified three-category version of the scale in which we put respondents into {*Left, Moderate, Right*} groups. We

present results from both models (for each country) below. In addition, and as before, we use the same modelling approach to analyse the results from our second experiment, based on the MFQ. Here, the correlation parameters capture the degree to which left- and right-wing respondents agree on the relative importance of the MFQ items when they think about morality.

Results

Figure 5 presents our estimates for each of the 74 MFV violations and the 30 MFQ items separately for those on the left and the right of the political spectrum from the model estimated using our three-category decomposition of ideology. The top and middle rows gives results from the UK and US versions of the MFV experiment, respectively, and the bottom row presents results from the MFQ experiment. While liberal and conservative rankings of abstract moral concerns are heterogeneous, those on the left and the right are largely uniform in their perceptions of the severity of concrete moral violations. Averaging across the five foundations, the correlation in the rankings of the MFV violations between liberals and conservatives is remarkably high in both the UK (0.94 [0.88, 0.97]) and the US (0.91 [0.84, 0.95]), and very few vignettes show differences between liberals and conservatives that are significantly different from zero. By contrast, the correlation in the rankings between left- and right-leaning respondents of the MFQ items is much lower (0.35 [-0.07, 0.69]), and for the sanctity and loyalty foundations the correlation is actually negative.

To investigate whether these results mask heterogeneity at more extreme ideological positions, we present the estimates for the correlation between respondents in each group ($\rho_{p,p'}$) of the seven-category ideological variable (alongside their associated 95% credibility intervals) in figure 6. The left and centre panels depict estimates from the MFV experiment in the UK and the US, and the right panel shows estimates from the MFQ experiment.

The figure shows that the correlation in perceptions of violation severity is positive for all pairs of ideological groups of respondents. The lowest correlation we measure is between those who are "Strongly left" and "Strongly right" in the UK, but even here the correlation is positive and reasonably
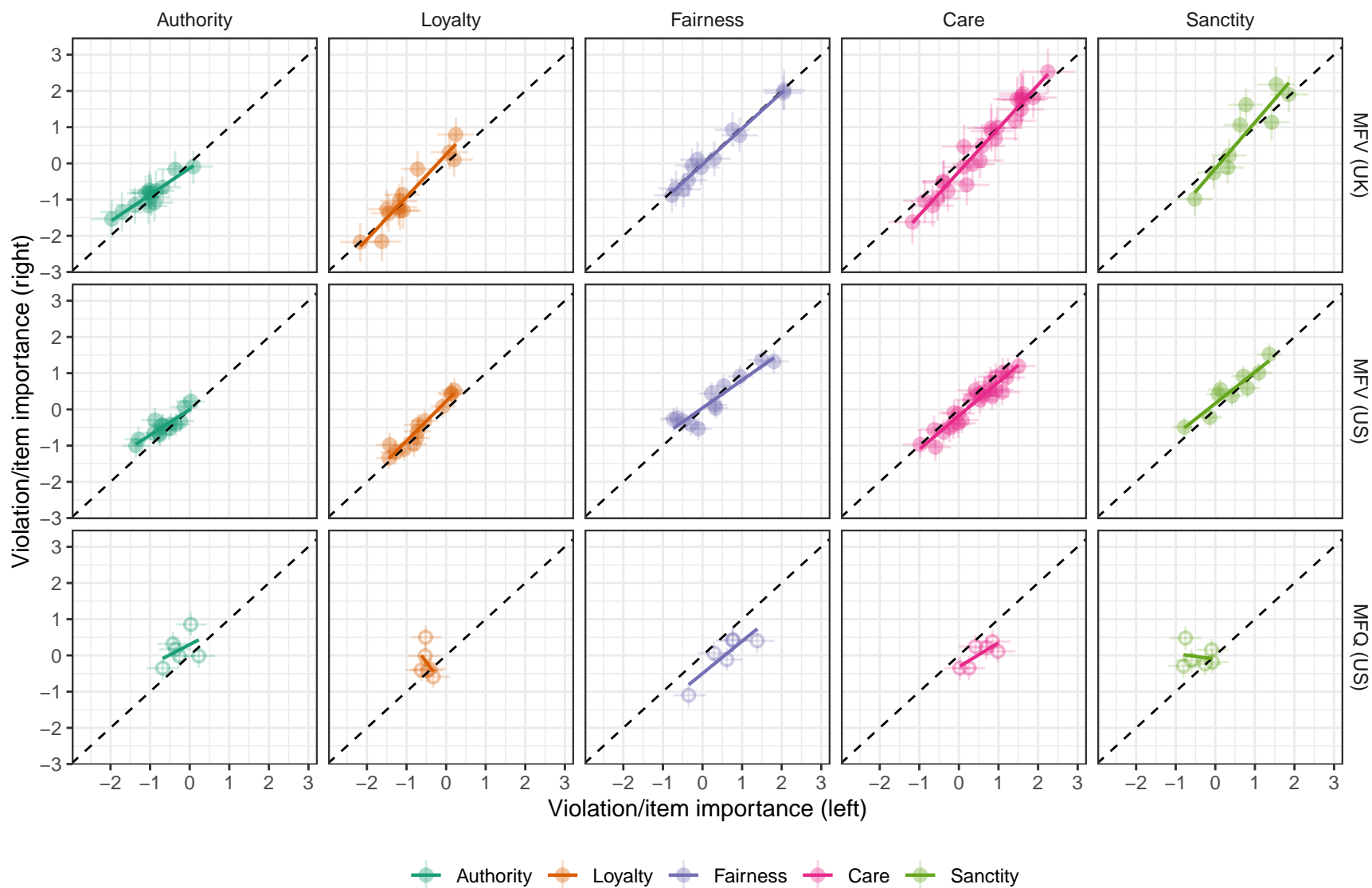
Figure 5: Correlation of violation severity/item importance between respondents on the left and right

strong at 0.61 [0.26, 0.87]. In the US, the correlation between "Strongly liberal" and "Strongly conservative" respondents is even higher at 0.82 [0.68, 0.91]. The respondents falling into these groups represent small fractions of the population, particularly in the UK. In the US, 23% of respondents describe themselves as being "Strongly" liberal or conservative, and in the UK just 4.5% report being "Strongly" left or right. It is therefore striking that even these small groups at the ideological extremes, regardless of their political differences, tend to have very similar moral intuitions about the relative severity of the violations that we included in our experiment.

By contrast, the right panel of figure 6 reveals that, when evaluating the relevance of the abstract moral statements contained in the MFQ, the rankings of items are very different for conservative and liberal respondents. For example, the estimated correlation between MFQ item rankings for "Strongly liberal" and "Strongly conservative" respondents is just 0.05 [-0.46, 0.57]. While we find that US liberals and conservatives largely agree on the relative severity of specific moral violations, there is almost no correlation in their judgments of the relevance of abstract moral principles. There is also a clear relationship between ideological proximity and the correlation of item strengths: the more ideologically proximate respondents are, the more they tend to agree about the relevance of different MFQ items. The stark difference between the highly correlated responses to the concrete moral violation vignettes and the low ideological correlation in responses to the MFQ items is hard to square with an interpretation that differences in moral *intuitions* between liberals and conservatives "illuminate the nature and intractability of moral disagreements" (Graham, Haidt and Nosek, 2009, 1029). Although liberals and conservatives may express different moral priorities when asked to reflect on their own moral reasoning, when confronted with concrete moral dilemmas they tend to react in strikingly similar ways.

Are the high correlations of perceived violation severity due to measurement error in our ideology variable? Our measure for political ideology is identical to the one used in existing work on MFT, but we can also re-estimate the correlated severity model to assess whether other voter-level characteristics are associated with different perceptions of violation severity. We find little evidence
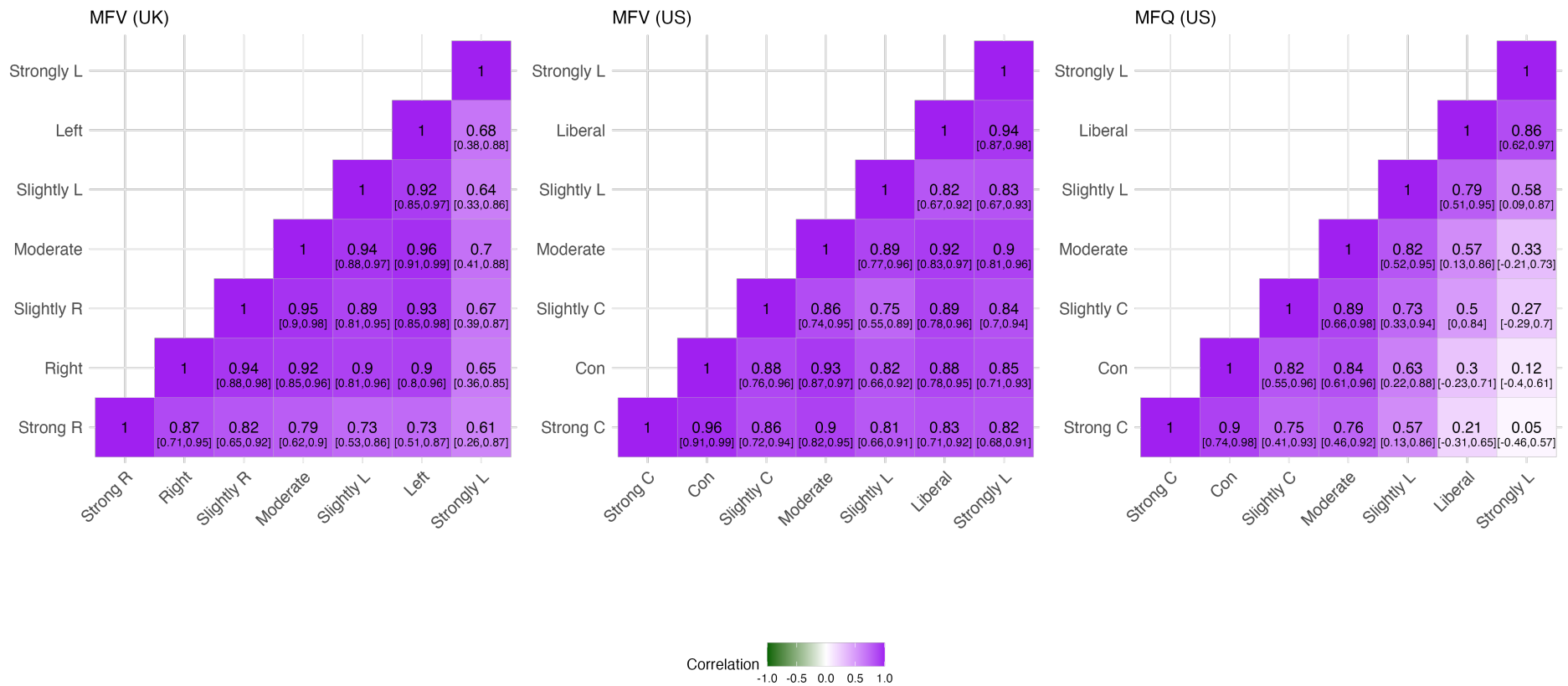
Figure 6: Correlation of violation severity by ideology

Note: Correlation in violation severity, across all foundations, for respondents with different self-reported ideological positions.

that this is the case. Most notably, the political voting histories of our respondents seem generally uninformative with respect to their moral intuitions. For instance, we find that, in the UK, the correlation in perceptions of the moral wrongness of the different violations is strongly positive between those who voted for the UK to "Leave" and those who voted to "Remain" in the EU referendum in 2016 (0.97), as well as between Conservative and Labour voters in the 2019 General Election (0.96). Likewise, Trump and Biden voters in the 2020 US Presidential Election also make very similar judgments about which violations are morally worse: the estimated ρ parameter for these groups is 0.95. In general, whether we use an attitudinal or behavioral measure of political ideology, the public consensus on relative violation severity explains far more variation in individual respondents' judgments of moral wrongness than do any systematic differences in moral judgment between respondents of different ideological or political groupings.

In the appendix, we use the same correlated strength model described above to analyse the similarities in moral judgment between respondents of different ages, education levels, genders, races and incomes. Looking across all the covariates we find a consistent pattern: the correlation in judgments of specific moral violations is always high, whereas the correlation in judgments of the relevance of the MFQ items is not. In general, while respondent demographics allow us to predict variation in the endorsement of abstract moral principles, there is very little disagreement across different groups of respondents when people make judgments of concrete moral violations. Finally, we find that respondents perceptions of violation severity are very similar regardless of whether the vignette describes the perpetrator of a particular transgression as male or female.

Conclusion

In this paper, we argued that existing empirical work on Moral Foundations Theory has over-stated differences in the moral intuitions of liberals and conservatives due to a mismatch between the theory and measurement. Measuring the moral priorities of individuals by asking them to explicitly reflect on the abstract principles of their own moralities is unlikely to capture the automatic and effortless

moral intuitions that lie at the conceptual heart of MFT. We proposed an alternative approach based on asking respondents to select between pairs of concrete moral transgressions, which comes closer to eliciting the types of intuitions that proponents of MFT claim underpin political disagreement between the left and the right. Empirically, we field two new survey experiments which show that while liberals and conservatives *articulate* different sets of moral priorities, we find that when confronted with specific moral comparisons they make very similar moral *judgments*. To the degree that political differences in moral evaluation do exist, these differences are small relative to the overall variation in judgments of different scenarios and small relative to the variation in support of abstract moral statements. As our final analysis demonstrates very clearly, while ideology strongly predicts the importance that citizens assign to abstract moral concerns, when making moral judgments of concrete scenarios related to those concerns citizens strongly agree with one another regardless of ideology.

Our findings have important implications for assessing potential explanations for contemporary political disagreement. In particular, a concern raised by previous studies on MFT is that the large moral differences between liberals and conservatives are likely to make the resolution of morally-loaded political issues intractable (Koleva et al., 2012; Haidt, 2012; Graham, Haidt and Nosek, 2009). Haidt (2012, 370-371) suggests that when disagreement is driven by instinctive moral responses, it becomes "difficult…to connect with those who live in other [moral] matrices". These fears are especially pronounced for "culture wars" issues – such as those related to sex, gender and multiculturalism – where voters' policy positions are particularly strongly associated with their expressed moral attitudes (Koleva et al., 2012). However, our results suggest this is unlikely to be the most important obstacle to productive conversation across political lines of disagreement. If conservatives and liberals react in largely similar ways to concrete moral questions (as we show in our MFV experiment), but express much more variation in their self-assessed moral attitudes (as documented in existing work and in our MFQ experiment), then the latter may reflect differences in how people talk about moral questions rather than genuine moral conflict. Indeed, our results align more with those who have argued for the unifying potential of morality in politics (Jung, 2022) and with work that suggests

that the real-world differences in moral behaviour between people with different politics "appear to be more a matter of nuance than stark contrast" Hofmann et al. (2014, 1342).

Looking beyond the literature on politics and morality, the central methodological argument we make – that survey questions asking people to reflect on abstract concepts can result in different response distributions than questions asking people to evaluate concrete manifestations of those concepts – has implications for many other literatures. For instance, much existing work in political behaviour draws on survey questions that ask respondents to reflect on their support for democracy and the normative ideals associated with democratic government. Would the ways in which responses to these questions correlate with respondent covariates persist if surveys instead asked respondents to evaluate specific violations of democratic ideals? Our findings raise the possibility that analyses based on soliciting reactions to specific stimulae might lead to very different patterns of "support for democracy" than findings based on questions that encourage voters to self-theorise about their democratic attitudes. In partial evidence of this point, recent survey-based work on the acceptability of political violence suggest that the specificity with which such questions are addressed can "cause the magnitude of the relationship between previously identified correlates and partisan violence to be overstated" (Westwood et al., 2022, 1).

One objection to the conclusion we draw is that the vignettes we use are mostly apolitical in nature, possibly suppressing political differences in moral expression. However, we think this property is helpful because it reduce endogenous responses where people infer the moral positions they think they *ought* to take on different issues as a result of their partisan or ideological allegiances. If we asked about a highly politicized moral issue – abortion, for instance – we might find highly polarized views between liberals and conservatives, but it would not be clear that such polarization stems from intuitive moral concerns or rather from the fact that voters have had their views on that issue deeply shaped by politics. In fact, our findings suggest that where prior work finds such political differences, they may not stem from fundamentally incompatible moral views on the importance of sanctity (or another foundation). More generally, proponents of MFT view moral intuitions as being

causally constitutive of political attitudes, but it is hard to see how the very high degree of consensus about moral judgments that we document could be the root cause of either policy-based or affective polarization between political groups. The differences in the intuitive morality of those on the left and right are simply too small to be responsible for the well-documented polarization between ideological groups.

References

Blumenau, Jack and Benjamin Lauderdale. 2022. "The Variable Persuasiveness of Political Rhetoric." *American Journal of Political Science* .

Bradley, Ralph Allan and Milton E Terry. 1952. "Rank analysis of incomplete block designs: I. The method of paired comparisons." *Biometrika* 39(3/4):324–345.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2017. "Stan: A probabilistic programming language." *Journal of statistical software* 76(1).

Ciuk, David J. 2018. "Assessing the contextual stability of moral foundations: Evidence from a survey experiment." *Research & Politics* 5(2):2053168018781748.

Clifford, Scott and Jennifer Jerit. 2013. "How words do the work of politics: Moral foundations theory and the debate over stem cell research." *The Journal of Politics* 75(3):659–671.

Clifford, Scott, Vijeth Iyengar, Roberto Cabeza and Walter Sinnott-Armstrong. 2015. "Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory." *Behavior research methods* 47(4):1178–1198.

Ditto, Peter H, David A Pizarro and David Tannenbaum. 2009. "Motivated moral reasoning." *Psychology of learning and motivation* 50:307–338.

Enke, Benjamin. 2020. "Moral values and voting." *Journal of Political Economy* 128(10):3679–3729.

Enke, Benjamin, Ricardo Rodríguez-Padilla and Florian Zimmermann. 2022. "Moral Universalism and the Structure of Ideology." *Working Paper* .

Eskine, Kendall J, Natalie A Kacinik and Jesse J Prinz. 2011. "A bad taste in the mouth: Gustatory disgust influences moral judgment." *Psychological science* 22(3):295–299.

Feinberg, Matthew and Robb Willer. 2015. "From gulf to bridge: When do moral arguments facilitate political influence?" *Personality and Social Psychology Bulletin* 41(12):1665–1681.

Feinberg, Matthew, Robb Willer, Olga Antonenko and Oliver P John. 2012. "Liberating reason from the passions: Overriding intuitionist moral judgments through emotion reappraisal." *Psychological science* 23(7):788–795.

Franks, Andrew S and Kyle C Scherr. 2015. "Using moral foundations to predict voting behavior: Regression models from the 2012 US presidential election." *Analyses of Social Issues and Public Policy* 15(1):213–232.

Frimer, Jeremy A, Jeremy C Biesanz, Lawrence J Walker and Callan W MacKinlay. 2013. "Liberals and conservatives rely on common moral foundations when making moral judgments about influential people." *Journal of personality and social psychology* 104(6):1040.

Garrett, Kristin N and Alexa Bankert. 2020. "The moral roots of partisan division: How moral conviction heightens affective polarization." *British Journal of Political Science* 50(2):621–640.

Gelman, Andrew and Iain Pardoe. 2006. "Bayesian measures of explained variance and pooling in multilevel (hierarchical) models." *Technometrics* 48(2):241–251.

Graham, Jesse, Brian A Nosek and Jonathan Haidt. 2012. "The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum." *PloS one* 7(12):e50092.

Graham, Jesse, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva and Peter H Ditto. 2011. "Mapping the moral domain." *Journal of personality and social psychology* 101(2):366.

Graham, Jesse, Jonathan Haidt and Brian A Nosek. 2009. "Liberals and conservatives rely on different sets of moral foundations." *Journal of personality and social psychology* 96(5):1029.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik and Peter H Ditto. 2013. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*. Vol. 47 Elsevier pp. 55–130.

Gray, Kurt and Jonathan E Keeney. 2015. "Impure or just weird? Scenario sampling bias raises questions about the foundation of morality." *Social Psychological and Personality Science* 6(8):859–868.

Grimmer, Justin and Christian Fong. 2021. "Causal Inference with Latent Treatments." *American Journal of Political Science* .

Haidt, Jonathan. 2001. "The emotional dog and its rational tail: a social intuitionist approach to moral judgment." *Psychological review* 108(4):814.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. London: Penguin.

Haidt, Jonathan and Craig Joseph. 2004. "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues." *Daedalus* 133(4):55–66.

Haidt, Jonathan, Craig Joseph et al. 2007. "The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules." *The innate mind* 3:367–391.

Haidt, Jonathan and Jesse Graham. 2007. "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize." *Social Justice Research* 20(1):98–116.

Hatemi, Peter K, Charles Crabtree and Kevin B Smith. 2019. "Ideology justifies morality: Political beliefs predict moral foundations." *American Journal of Political Science* 63(4):788–806.

Hewitt, Luke B. and Ben M. Tappin. 2022. "Rank-Heterogeneous Effects of Political Messages: Evidence from Randomized Survey Experiments Testing 59 Video Treatments." *Working Paper* .

Hofmann, Wilhelm, Daniel C Wisneski, Mark J Brandt and Linda J Skitka. 2014. "Morality in everyday life." *Science* 345(6202):1340–1343.

Iyer, Ravi, Spassena Koleva, Jesse Graham, Peter Ditto and Jonathan Haidt. 2012. "Understanding libertarian morality: The psychological dispositions of self-identified libertarians.".

Jost, John T. 2012. "Left and right, right and wrong.".

Jung, Jae-Hee. 2020. "The mobilizing effect of parties' moral rhetoric." *American Journal of Political Science* 64(2):341–355.

Jung, Jae-Hee. 2022. "Voters' Preferences for Parties' Moral Rhetoric." *Working Paper* .

Kalla, Joshua L, Adam Seth Levine and David E Broockman. 2022. "Personalizing moral reframing in interpersonal conversation: A field experiment." *The Journal of Politics* 84(2):1239–1243.

Kertzer, Joshua D, Kathleen E Powers, Brian C Rathbun and Ravi Iyer. 2014. "Moral support: How moral values shape foreign policy attitudes." *The Journal of Politics* 76(3):825–840.

Koleva, Spassena P, Jesse Graham, Ravi Iyer, Peter H Ditto and Jonathan Haidt. 2012. "Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes." *Journal of research in personality* 46(2):184–194.

Milesi, Patrizia. 2016. "Moral foundations and political attitudes: The moderating role of political sophistication." *International Journal of Psychology* 51(4):252–260.

Rao, PV and Lawrence L Kupper. 1967. "Ties in paired-comparison experiments: A generalization of the Bradley-Terry model." *Journal of the American Statistical Association* 62(317):194–204.

Rempala, Daniel M, Bradley M Okdie and Kilian J Garvey. 2016. "Articulating ideology: How liberals and conservatives justify political affiliations using morality-based explanations." *Motivation and Emotion* 40(5):703–719.

Schnall, Simone, Jonathan Haidt, Gerald L Clore and Alexander H Jordan. 2008. "Disgust as embodied moral judgment." *Personality and social psychology bulletin* 34(8):1096–1109.

Simonsen, Kristina B and Bart Bonikowski. 2022. "Moralizing Immigration: Political Framing, Moral Conviction, and Polarization in the United States and Denmark." *Comparative Political Studies* p. 00104140211060284.

Smith, Kevin B, John R Alford, John R Hibbing, Nicholas G Martin and Peter K Hatemi. 2017. "Intuitive ethics and political orientations: Testing moral foundations as a theory of political ideology." *American Journal of Political Science* 61(2):424–437.

Suhler, Christopher L and Patricia Churchland. 2011. "Can innate, modular "foundations" explain morality? Challenges for Haidt's moral foundations theory." *Journal of cognitive neuroscience* 23(9):2103–2116.

Talaifar, Sanaz and William B Swann Jr. 2019. "Deep alignment with country shrinks the moral gap between conservatives and liberals." *Political Psychology* 40(3):657–675.

Walter, Annemarie S and David P Redlawsk. 2021. "The Effects of Politician's Moral Violations on Voters' Moral Emotions." *Political Behavior* pp. 1–27.

Westwood, Sean J, Justin Grimmer, Matthew Tyler and Clayton Nall. 2022. "Current research overstates American support for political violence." *Proceedings of the National Academy of Sciences* 119(12):e2116870119.

Wheatley, Thalia and Jonathan Haidt. 2005. "Hypnotic disgust makes moral judgments more severe." *Psychological science* 16(10):780–784.