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

Jack Blumenau *University College London*
Benjamin E Lauderdale *University College London*

Moral Foundations Theory (MFT) aims to explain the origins of and variation in human moral reasoning. Applications 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. However, differences in self-reported assessments of the moral relevance of each foundation do not imply that citizens with different political orientations respond to concrete scenarios based on different moral intuitions. We introduce a new approach for measuring the implicit importance of the 5 moral foundations by asking survey respondents from the UK and the US to compare pairs of vignettes which describe 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.

A core concern in contemporary democratic politics is whether adherents to different political ideologies might have fundamentally incompatible moral outlooks. Debates between conservatives and liberals on issues as disparate as abortion, gun control, LGBTQ+ rights, immigration, and climate change are often marked by highly-charged moral rhetoric, with participants on both sides appearing to be intransigent and unwilling to compromise. Indeed, scholars have speculated that at the core of increasing disagreement and hostility between political groups in advanced democracies (Abramowitz and Saunders, 2008; Iyengar et al., 2019; Hobolt, Leeper and Tilley, 2021) are a set of contrasting moral intuitions that divide political opponents (Haidt, 2012; Koleva et al., 2012).

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

[*] **This version**: June 15, 2022

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. While liberals tend to prioritize the "individualizing" foundations of care and fairness, conservatives put weight on all five foundations, including the "binding" foundations of authority, loyalty, and sanctity (Graham, Haidt and Nosek, 2009; Graham et al., 2011; Haidt and Graham, 2007; Kertzer et al., 2014; Koleva et al., 2012). On the basis of these findings, proponents have concluded that there is an "invisible wall separating liberal and conservative moralities" Haidt and Graham (2007, 111). This message – that liberals and conservatives live in partitioned moral worlds – has been promoted far beyond the confines of academia,[1] to the point where the idea of a yawning moral divide between the left and right appears to have been accepted as self-evident.[2]

However, 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

[1] Jonathan Haidt's TedTalk on the subject, for example, has been viewed over 4 million times, and his book summarizing this research agenda (Haidt, 2012) reached the New York Times best-seller list in 2012.

[2] See, for example, The Moral Chasm That Has Opened Up Between Left and Right Is Widening, New York Times, October 27, 2021

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. This approach to measuring moral priorities is in tension, however, with a core theoretical assumption of MFT. Though the lens of MFT, moral judgments are thought to be made via automatic processes in which people do not have any conscious awareness of the factors that lead to a particular conclusion being reached (Haidt, 2001, 1029). The MFQ, by contrast, prompts respondents to self-assess the motivations for their moral choices; motivations which are – by MFT's own assumptions – inaccessible to them. We argue that by prompting respondents to provide self-theories of their own morality, rather than soliciting reactions to specific moral stimulae, the MFQ is likely to overestimate political differences, both because it increases the opportunity for motivated moral reasoning (Ditto, Pizarro and Tannenbaum, 2009) and because liberals and conservatives are likely to imagine different moral scenarios when confronted with abstract descriptions of the content of each foundation.

Second, 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. While this is a common approach across many areas of research into public opinion and political psychology, it raises questions regarding both the internal (Grimmer and Fong, 2021) and external (Blumenau and Lauderdale, 2022) validity of the findings. While liberal and conservative respondents may give different reactions to a few commonly-used vignettes describing moral violations, we are interested in whether they have different intuitions *on average* when presented with the range of moral transgressions relevant to each foundation.

Third, 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, as they do not force respondents to make direct comparisons between foundations. Because the MFQ asks respondents to rate items relevant to each foundation individually, rather than compare items from different foundations, it is likely to overstate political differences because respondents are not forced consider the moral foundations on a common, comparable scale. They can express greater or less moral concern about individual foundations without the constraints on consistency that come from comparisons of concrete scenarios. 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 respondents to simply compare the vignettes, and then indicate which they think constitutes the worse moral transgression. Our treatment is therefore designed to elicit the intuitive moral responses that are central to MFT (Haidt, 2001), and the paired-comparison design also means that we recover direct information about how respondents trade-off different kinds of violations and their underlying foundations. Finally, we provide a large number of implementations of violations for each foundation (drawn from those described in Clifford et al. (2015)), which enables us to be more confident that the results we present are not attributable to idiosyncratic differences that arise from any given moral transgression vignette.

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. In particular, we show that in our set of moral violations, sanctity considerations appear to be significantly more important for informing the implicit moral responses that we capture than they are in the MFQ when respondents provide more explicit evaluations of their moral priorities. This could be because of the particular violations we happened to present, but this concern about the representativeness of the violations statements also applies to existing survey evidence relating to moral foundations theory, which calls into question how much we really know about the absolute levels at which the different foundations are weighted in individuals' moral assessments.

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. When making moral judgments, conservatives do put more weight on authority and loyalty considerations than liberals, and liberals prioritize harm and fairness concerns more than conservatives, but these differences are small relative to the average differences between the foundations and they are very small relative to the average differences between specific moral violation scenarios. 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. The key implication of this result is that, at least when assessed through their intuitive responses to concrete violations of moral principles, there is far more that unites the moralities of liberals and conservatives than there is that divides them.

We see our design as providing a sharp test of the 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), 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. Indeed, our experiment is designed explicitly to solicit the types of fast, automatic, intuitive judgments that are central to MFT. 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.

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.[3]

Understanding the moral decisions that people make therefore requires understanding the sources 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 faced by humans throughout history: the need to protect the vulnerable (and 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; fair-

[3] In contrast to rationalist perspectives which view conscious reasoning as the main determinant of moral decision-making (Kohlberg, Levine and Hewer, 1983; Kolberg, 1984), MFT therefore views moral reasoning as no different to other forms of reasoning in that it is likely to be motivated (Kunda, 1990; Ditto, Pizarro and Tannenbaum, 2009).

ness; loyalty; authority; and sanctity – which are the “irreducible basic elements” (Graham et al., 2013, 56) needed to explain and understand the moral domain.[4] 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 refereed to as the *binding* foundations (because they emphasize virtues connected to binding individuals together into well-functioning groups).[5]

The 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 endorsing the foundations to different degrees, and therefore developing distinct moral worldviews. As a consequence, while some societies, or groups, will construct moralities on the basis of only one or two of the foundations, others will adopt broader views of moral and immoral behavior.

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 are shaped by experience – in order 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

[4] Notably, the five foundation structure suggested by MFT constitutes a significantly wider moral domain than that envisaged by other authors (e.g. Turiel, 1983), who tend to restrict discussions of morality to considerations of either harm or fairness.

[5] In more recent iterations of the theory, Iyer et al. (2012) propose a sixth – liberty/oppression – foundation, which captures the impulse for people to oppose those who dominate them and restrict their liberty. 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.

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' political ideology to questions designed to measure reliance on each of the foundations in moral decision-making, appears to offer support for this hypothesis.[6] 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). 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 "Conser-

[6] Aside from surveys, a second source of evidence for the idea that there are pronounced moral differences between liberals and conservatives comes from the propensity for liberals and conservatives to use different types of moral argument at different rates. Graham, Haidt and Nosek (2009), for example, show that sermons delivered by pastors in liberal churches contain more appeals to care and fairness considerations, and fewer references to authority and sanctity concerns, than sermons delivered in conservative churches. When asked to make a persuasive argument on a political issue, liberal US survey respondents tend to make arguments that are deploy the "individualizing" foundations while conservatives are more likely to make use of the "binding" foundations in their arguments (Feinberg and Willer, 2015). Similarly, when asked to justify their their party affiliations, US undergraduates also tend to give arguments that correspond to the political predictions of MFT (Rempala, Okdie and Garvey, 2016). Differences in use of words relevant to the foundations is also apparent in liberal and conservative newspaper reporting of political issues (Clifford and Jerit, 2013), as well as in speeches made in the US senate (Sagi and Dehghani, 2014). There is also evidence that liberals are somewhat more responsive to arguments based on the individualizing foundations, and less responsive to "binding" arguments, than are conservatives (Day et al., 2014).

vatives have many moral concerns that liberals simply do not recognize as 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).[7] 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.[8]

[7] Koleva et al. (2012, 184), for instance, argue that "moral intuitions are one powerful and largely unexplored psychological mechanism that underlies ideology in general and issue positions in particular."

[8] One exception to this consensus is that, at least when judging the character of influential figures in history, liberals and conservatives appear to rely on similar sets of moral foundations (Frimer et al., 2013). However, the findings described in Frimer et al. (2013) relate primarily to a sample of American college professors (and an even smaller, non-representative sample of MTurk participants) who might be expected to have similar moral outlooks that are less sensitive to their

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.[9] 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 specific moral statements. For instance, respondents signal the strength of their agreement for 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).

Despite the ubiquity of this survey instrument in existing work, 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,

political orientations.

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

the vast majority 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. 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 a survey item such as "Whether or not someone showed a lack of respect for authority", people might imagine very different scenarios, which may differ in moral importance. For instance, a liberal might imagine a situation in which a child is rude to their parent, while a conservative might imagine a situation in which a soldier refuses to follow the instructions of their commanding officer. While the liberal might not think that the child has committed a moral offence, the conservative is likely to think the soldier has done something wrong. As a consequence, we might expect the liberal and conservative to give differing responses about the importance of authority to their own moral codes. However, such responses would not reflect differences in the intuitive moral judgments of the liberal and the conservative, but rather differences in the specific scenarios that they associate with the abstract statement. If the liberal was to be asked about the soldier's behaviour, or the conservative about the child's, then we might imagine high levels of agreement between the two groups about the relative severity of the two scenarios. 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 these scenarios may differ dramatically 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, though this is a key explanatory goal of MFT. In particular, the MFQ contains sets of items which respondents are asked to rate one at a time, without providing direct comparisons between items relating to different foundations. When considering the importance of a given moral category in isolation, respondents may deem it to play an important role in their moral calculus, but when presented in competition with another moral category it may seem less important. A respondent might think, for instance, that "whether or not someone showed a lack of loyalty" is a relevant moral concern, but the importance assigned to that statement could decrease substantially when compared directly to "whether or not someone used violence". 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). In addition to the fact that comparison designs tend to be more successful, relative to single-rating designs, at generating survey estimates that replicate real-world behavioral benchmarks (Hainmueller, Hangartner and Yamamoto, 2015), pitting moral foundations against one another in paired competition would take seriously the idea that moral values tend to work in a "competitive and cooperative" fashion (Ciuk and Jacoby, 2015, 709).

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). Reliance on a small number of issues might lead to over-estimates 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."

A further difficulty faced by 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 is that they are subject to potential confounding concerns, as 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). In this context, as Gray and Keeney (2015, 859) suggest, 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". 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.

Before proceeding, we note one study that shares greater similarity with our own. In study 2 of his PhD dissertation (Graham, 2010), Jesse Graham, like us, presents an analysis of a survey experiment which presents respondents with paired comparisons of moral violations. Our approach makes a number of important advances on Graham's work. First, Graham tests three examples of moral violations per foundation, whereas we test about five times as many examples of violations. Second, the violations he tests are somewhere between explicit moral self-assessments of the MFQ and concrete scenarios that would reveal implicit moral evaluations. The three "harm" violations in his design, for instance, are "Doing something cruel", "Acting harmfully", and "Hurting someone's feelings", all of which are abstract and imprecise regarding severity. In contrast, we use a set of treatment texts that describe concrete instances of a particular foundation being violated, making it more likely that we will elicit intuitive moral responses from our survey respondents. Third, we use larger, representative samples from two countries. Fourth, we show below how we can combine this design with a multilevel model which allows us to quantify the average and distribution of effects for violations each foundation, and the degree to which these effects differ as a function of political ideology.

Experimental Design

Moral Foundations Violations

We base our design 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. Importantly, 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.[10]

We have lightly edited these vignettes for use in our context. In particular, as we field these vignettes 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.[11] 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.[12] 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.[13] The sample of violations is not balanced 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 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

[10] The analysis in Clifford et al. (2015) is based on 510 responses to a non-representative survey in which respondents rated the moral wrongness of over 100 vignettes, completed the MFQ, and answered a series of political questions. In addition to using two large and nationally representative samples, and a significantly simpler and shorter 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. As argued above, comparative evaluations are important for prompting respondents to consider the trade-offs between foundations.

[11] For example, "A woman leaving her dog outside in the rain after it dug in the **trash**" is converted to "A woman leaving her dog outside in the rain after it dug in the **rubbish**" in the UK implementation of the treatment.

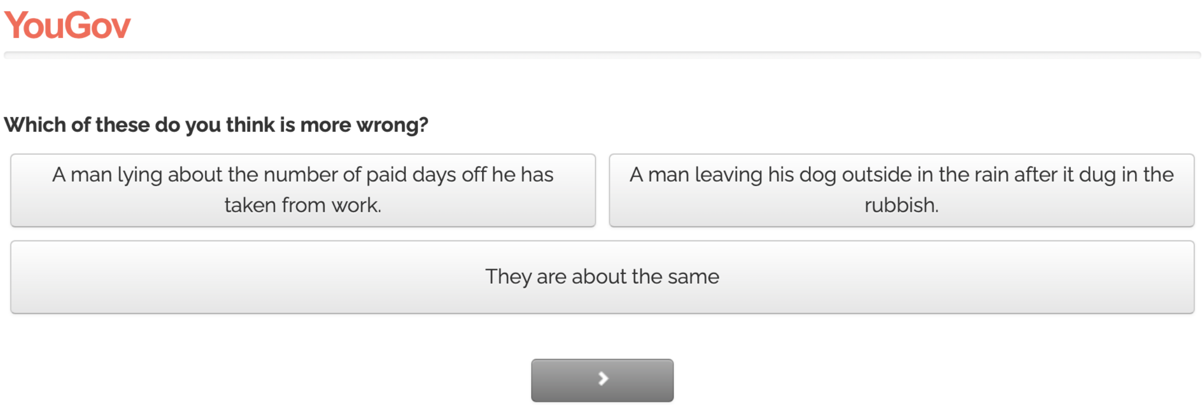
[12] For example, "A **woman** leaving **her** dog outside in the rain after it dug in the trash" is converted to "A **man** leaving **his** dog outside in the rain after it dug in the trash" in the male implementation of the treatment.

[13] We also excluded one vignette about having intimate relations with a recently deceased loved one because we thought it was likely to push the limits of what constitutes an acceptable survey question (even for YouGov respondents).

by examining their judgments of these scenarios. As argued above, 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 (Graham, Haidt and Nosek, 2009; Haidt, 2001).

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 in this application 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, and increase our confidence that any differences between liberals and conservatives that we *do* detect are attributable to moral intuitions relevant to each foundation.

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 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.



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: Experiment prompt.

Randomization, Prompt and Sample

We present random pairs of the vignettes described above to survey respondents and ask them to select which of a given pair of violations is "more wrong".[14] 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 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".

For each comparison, we first sampled two foundations and then, conditional on the foundation drawn we sampled two of the vignettes, without replacement, from the full set of 74 violations. 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 vignette, we also randomly sampled whether the person committing the relevant violation was a man or a woman. Each respondent answered 6 pairwise

[14] Before completing the paired comparison task, respondents first read an information screen which explained that they were about to answer questions which were part of a piece of research "on public views about morality". Respondents were also warned about the content of some of the comparisons: "Some of the questions on this survey include tasks that involves making judgements on situations that people may find objectionable or immoral". Respondents were also provided with the option to opt-out of the survey. Our study received ethical review before it was fielded by the UCL Research Ethics panel (project id: 21793/001).

comparisons and 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.[15]

Measuring Moral Intuitions

Model definition

Our experiment results in an ordered response variable with three categories:

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

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:

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

where θ_k is the cutpoint for response category k. We can interpret the α_j as the "severity" of violation j. This parameter is increasing in the frequency with which our respondents choose violation j as the "worse" moral violation in paired-comparison with other violations. α_j will also be larger when the violations that violation j "defeats" are themselves more severe, and will be smaller when violation j only defeats other less severe 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

[15] 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. In the UK, 1511 completed all six comparisons, 70 completed 5, 11 completed 4, and 6 completed between 1 and 3 comparisons. In the US, 2236 completed all six comparisons, 107 completed 5, 20 completed 4, and 12 completed between 1 and 3 comparisons.

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 is a linear combination of the average severity of violations of a given foundation type and the severity 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. Note also that, given the way that the $\mu_{f(j)}$ enter into equation 2, they cancel in the instances where violation j and j' are of the same foundation type, meaning that – in those instances – the probability that one violation is "worse" than another is determined solely by the difference in the violation-specific random effects, ν_j.

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. Note that 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 con-

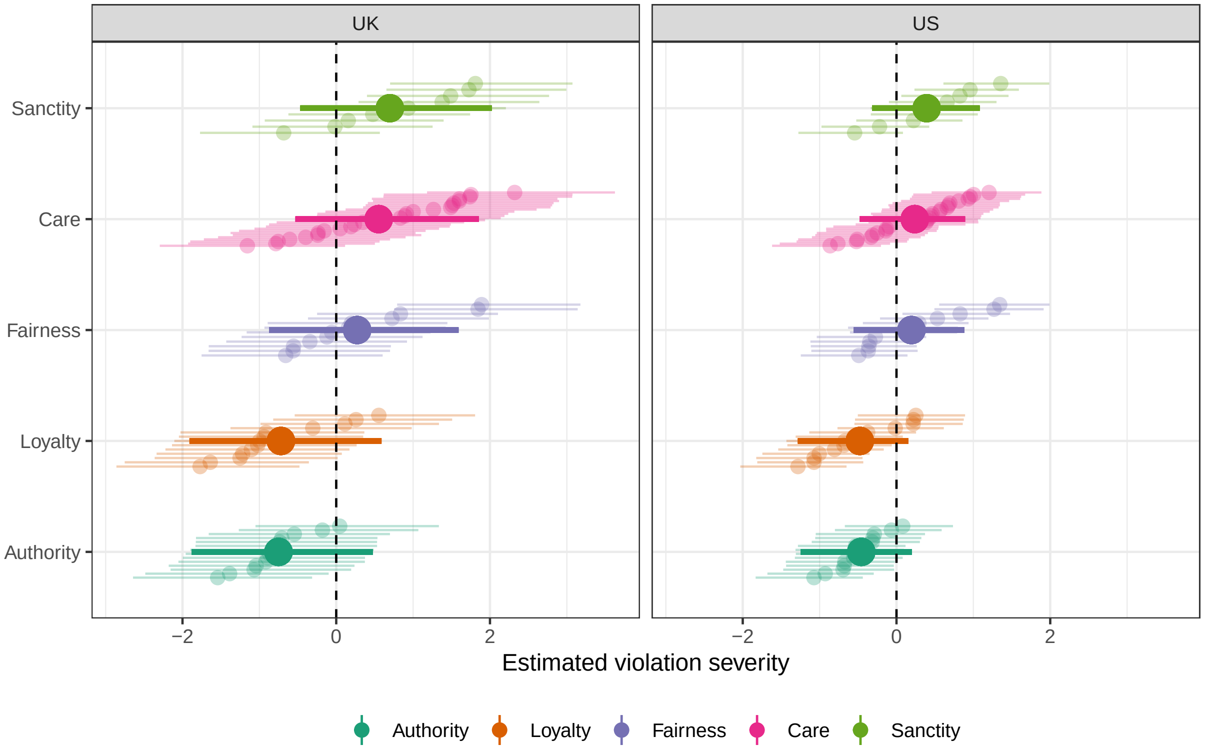


Figure 2: Estimates of $\mu_{f(j)}$ and α_j from equations 2 and 3.

venience samples that are unlikely to be representative. We maximize the representativeness of our estimates 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 2. The figure shows the estimated average severity of each of the five moral foundations (μ_f) in both the UK (left panel) and the US (right panel). It also depicts the estimated severity of each of the 74 individual moral violations ($\mu_{f(j)} + \nu_j$) that we include in the experiment (transparent points and intervals).

The figure reveals two main findings. First, we recover systematic differences in the average severity of the tested violations across the different foundations. 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.

This finding is mostly consistent with existing work. 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, in contrast to previous work, we find that our 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. This contrasts with findings presented in Graham, Haidt and Nosek (2009), where – even among the most conservative respondents – sanctity considerations are considered less important to moral decision-making than either harm or fairness concerns. This difference is potentially a consequence of the different design choices we make versus those made by Graham, Haidt and Nosek (2009). 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.

It could also be that we selected sanctity violations that were 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. As noted above, and returned to below, it is difficult to define a population of violations in a way that would allow one to ensure representativeness. 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.

This observation leads to our second main finding of this initial model, that we observe significant heterogeneity in violation severity within each foundation type. In both the UK and the US, the average effects for care, fairness and severity indicate that violations of these types are worse than those of other foundations, but some violations of these foundations are significantly weaker. Moreover, there is a large degree of overlap in the severity of violations of different types. This suggests that only a small fraction of the variation in respondents' evaluations of severity can be explained by the foundations to which the violations apply. Calculating the proportion of variance in the α_j parameters that is explained by the foundation predictors in equation 3, we find that the foundation effects ($\mu_{f(j)}$) explain 37% of the variation in violation severity for UK respondents and 31% of the variation for US respondents.[16] Consequently, though the moral judgments of our respondents for individual vignettes is clearly predicted by which foundation the vignettes were associated with, approximately 2/3rds of the variation in violation severity remains unexplained in this model.

We present the 4 violations considered most and least severe for each foundation, in each country, in tables 1 and 2. Qualitative inspection of these tables reveal some interesting patterns about the sources of within-foundation variation in violation severity. For instance, the most severe instances of violations of the care foundation tend to contain examples of physical harm (hitting a child; whipping a pony; attacking cats), while care violations that focus on emotional harm (e.g. criticizing someone's work; casting doubt on a person's appearance) tend to be perceived as less severe. That our respondents appear to make a distinction between between physical and emotional forms of harm is consistent with other research showing that moral violations involving bodily harm elicit distinct reactions (Heekeren et al., 2005; Clifford et al., 2015). Similarly, respondents appear to object more to scenarios in which a person demonstrates a lack of loyalty to their country than to scenarios in which they exhibit a lack of loyalty towards their family, school, or employer. Finally, it is also notable that 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 (a point that we return to below).

[16] 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.

Table 1: Most severe violations of each foundation, by country

UK	US
Care	
A girl/boy setting a series of traps to kill stray cats in her/his neighborhood.	A girl/boy setting a series of traps to kill stray cats in her/his neighborhood.
A woman/man laughing at a disabled co-worker while at an office party.	A woman/man laughing at a disabled co-worker while at an office party.
A woman/man throwing her/his cat across the room for scratching the furniture.	A woman/man throwing her/his cat across the room for scratching the furniture.
A woman/man lashing her/his pony with a whip for breaking loose from its pen.	A woman/man laughing as she/he passes by a cancer patient with a bald head.
Fairness	
A judge taking on a criminal case although she/he is friends with the defendant.	A judge taking on a criminal case although she/he is friends with the defendant.
A politician using public money to build an extension on her/his home.	A politician using federal tax dollars to build an extension on her/his home.
A professor giving a bad grade to a student just because she/he dislikes him.	A professor giving a bad grade to a student just because she/he dislikes him.
A referee intentionally making bad decisions that help her/his favoured team win.	A referee intentionally making bad calls that help her/his favoured team win.
Authority	
An employee trying to undermine all of her/his boss' ideas in front of others.	A girl/boy ignoring her/his father's orders by taking the car after her/his curfew.
A girl/boy ignoring her/his father's orders by taking the car after her/his curfew.	An employee trying to undermine all of her/his boss' ideas in front of others.
A woman/man talking loudly and interrupting the mayor's speech to the public.	A woman/man talking loudly and interrupting the mayor's speech to the public.
A woman/man turning her/his back and walking away while her/his boss questions her/his work.	A woman/man turning her/his back and walking away while her/his boss questions her/his work.
Loyalty	
A British woman/man telling foreigners that the UK is an evil force in the world.	The US Ambassador joking in Great Britain about how stupid she/he thinks Americans are.
The UK Ambassador joking in America about how stupid she/he thinks the British are.	An American woman/man telling foreigners that the US is an evil force in the world.
A British film star saying she/he agrees with a foreign dictator's denunciation of the UK.	An American movie star saying she/he agrees with a foreign dictator's denunciation of the US.
A former UK Army General saying publicly she/he would never buy any UK product.	A former US General saying publicly she/he would never buy any American product.
Sanctity	
A woman/man in a bar using her/his phone to watch people having sex with animals.	A small group of religious women/men eating the flesh of their naturally deceased members.
A small group of religious women/men eating the flesh of their naturally deceased members.	A woman/man in a bar using her/his phone to watch people having sex with animals.
An employee at a morgue eating her/his pepperoni pizza off of a dead body.	A woman/man having sex with a frozen chicken before cooking it for dinner.
A woman/man having sex with a frozen chicken before cooking it for dinner.	An employee at a morgue eating her/his pepperoni pizza off of a dead body.

Table 2: Least severe violations of each foundation, by country

UK	US
Care	
A girl/boy telling a boy that his older brother is much more attractive than him.	A girl/boy making fun of her/his brother for getting dumped by his girl/boyfriend.
A woman/man telling a man that his painting looks like it was done by children.	A girl/boy telling a boy that his older brother is much more attractive than him.
A girl/boy making fun of her/his brother for getting dumped by his girl/boyfriend.	A woman/man telling a man that his painting looks like it was done by children.
A woman/man quickly canceling a blind date as soon as she/he sees the man.	A girl/boy telling her/his classmate that she/he looks like she/he has gained weight.
Fairness	
A girl/boy skipping to the front of the queue because her/his friend is an employee.	A woman/man lying about the number of vacation days she/he has taken at work.
A woman/man playing football pretending to be seriously fouled by an opposing player.	A girl/boy skipping to the front of the line because her/his friend is an employee.
A woman/man cheating in a card game while playing with a group of strangers.	A woman/man playing soccer pretending to be seriously fouled by an opposing player.
A tenant bribing her/his landlord to be the first to get her/his flat repainted.	A woman/man cheating in a card game while playing with a group of strangers.
Authority	
A woman/man secretly watching sport on her/his mobile phone during church.	A woman/man secretly watching sports on her/his cell phone during church.
An intern disobeying an order to dress professionally and comb her/his hair.	An intern disobeying an order to dress professionally and comb her/his hair.
A player publicly yelling at her/his football coach during a game.	A student stating that her/his professor is a fool during an afternoon class.
A star player ignoring her/his coach's order to come off the pitch during a game.	A player publicly yelling at her/his soccer coach during a game.
Loyalty	
The head girl/boy saying that her/his rival school is a better school.	The class president saying that her/his rival college is a better school.
A woman/man secretly voting against her/his husband in a local talent competition.	A woman/man secretly voting against her/his husband in a local talent competition.
A former politician publicly giving up her/his citizenship to the UK.	An employee joking with competitors about how badly her/his company did last year.
A teacher publicly saying she/he hopes another school wins a competition.	A local politician saying that the neighboring town is much better than her/his town.
Sanctity	
A single woman/man ordering an inflatable sex doll that looks like her/his assistant.	A single woman/man ordering an inflatable sex doll that looks like her/his assistant.
A woman/man marrying her/his first cousin in an elaborate wedding.	A woman/man searching through the trash to find men's discarded underwear.
A woman/man searching through the rubbish to find men's discarded underwear.	A woman/man in a bar offering sex to anyone who buys her/him a drink.
A woman/man in a bar offering sex to anyone who buys her/him a drink.	A drunk elderly woman/man offering to have oral sex with anyone in the bar.

Measuring Moral Intuitions by Ideology

Model definition

The model described above allows us to characterize the degree to which the foundation relevant to a moral violation determines judgments of the severity of that violation. 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 allow us to describe how these foundation-level effects vary by respondent ideology.

To evaluate the ideological hypothesis, we therefore modify the model in equation 2 to allow the violation-level parameters, α_j, to vary according to the self-reported ideological position of the respondent. We follow Graham, Haidt and Nosek (2009) and ask respondents to place themselves on a seven-point ideological scale before they complete the violation comparison task.[17] 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 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

[17] Respondents in the US select their position 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*}. This question also allowed respondents to provide a "Don't know" response, which was selected by 316 individuals in our UK sample and 200 individuals in our US sample. To avoid dropping these respondents from the analysis, we recode them as "moderate" on our ideology variables. None of the results we present are sensitive to this choice.

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, but still allows for large deviations from one group to the next if the information from the data is sufficiently strong. In essence, this set-up assumes that the moral intuitions of those who identify as "Strongly liberal" are likely to be similar to those who identify as "Liberal", and allows the model to partially pool the interaction-effects of those groups towards each other.

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 3. The horizontal dashed lines represent the estimates of violation severity for each foundation for the moderate ideological group, and the point-estimates and 95% credibility intervals are provided for all other groups.

The figure reveals that some, but by no means all, of the predictions of moral foundations theory are supported in this data. For instance, we find that – as expected – respondents on the right of the political spectrum perceive violations of the loyalty foundation as somewhat more important,

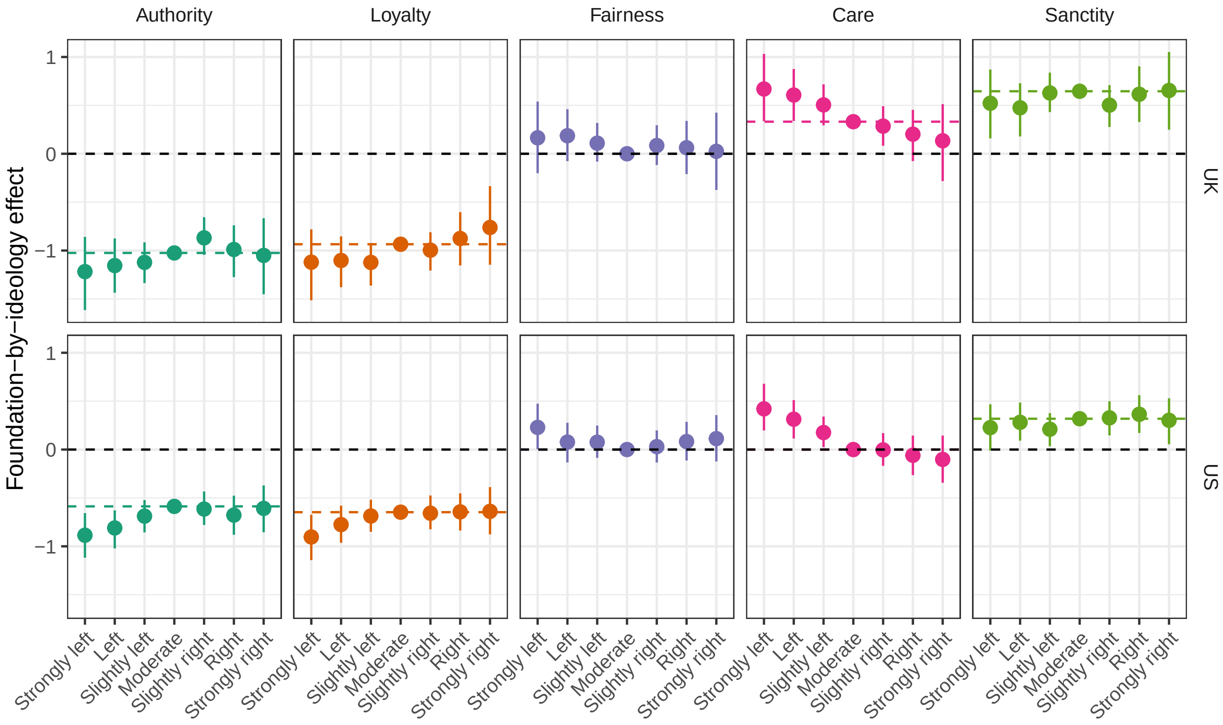


Figure 3: Estimates of $\mu_f + \gamma_{f,p}$ from equation 5.

and violations of the care foundation as somewhat less important, than respondents on the left of politics. These patterns hold for both the UK and US samples, and the ideological gradient is especially pronounced for the care foundation. There is also evidence that, again consistent with MFT, left-wing respondents care less about authority violations that right-wing respondents.

However, we find little evidence in support of the predicted relationship between fairness concerns and ideology, nor between sanctity concerns and ideology. 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 essentially no difference at all 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.

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 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. While it is not unusual for interaction effects to be smaller than the main effects of a treatment, proponents of MFT do clearly believe that such interactions are critical determinants of moral judgment. As Graham, Haidt and Nosek (2009, 1033) suggest, the "moral thinking of liberals and conservatives may not be a matter of more versus less but of different opinions about what considerations are relevant to moral judgment." Our results, by contrast, suggest that liberals and conservatives largely react to the same considerations when making moral judgments: their choices reflect broadly similar orderings of the importance of the five foundations, albeit with marginally different emphases.

An important implication of these results is that estimates of political differences in moral judgment depend heavily on whether survey items aim to capture explicitly stated moral principles versus implicit moral responses to concrete situations. When asked to articulate theories of their own morality, conservatives cite 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 a scenario in which a person has sex with a chicken.

Measuring Agreement about Individual Violation Severity

Model definition

In the previous section, we were interested in exploring whether there were extensive political differences between liberals and conservatives in the judgments they make, on average, of violations relevant to each foundation. Given the small differences 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 assessments of specific moral violations, even if these differences do not map neatly onto the foundation-based categorization proposed by MFT.

One way of assessing the correlation in judgments of violation wrongness would be to generate estimates of severity for each of our 74 vignettes separately for liberals and conservatives, and then to calculate the correlation between those estimates. More positive correlations imply that, on average, different groups of respondents agree about which violations are more wrong, while less positive or negative correlations would imply that different groups of respondents view different violations as more wrong. However, because we have a relatively small number of observations for each violation, such an approach would likely lead to us underestimate the true correlation in violation importance for respondents with different self-reported ideological positions, simply due to the imprecision of our estimates for each violation in each ideological sub-population. Instead, we adopt a modelling framework in which we directly estimate the correlation of moral judgments between people with different ideological stances.

We build on the first-stage model described in equation 4 in order to determine whether different groups of people rate violations as worse than others in different ways on average. We then 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 experiment, tend to be very similar for pair of 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.

Results

Figure 4 presents our estimates of violation severity for each of the 74 vignettes separately for those on the left and the right of the political spectrum from the model estimated using our three-category decomposition of ideology. Averaging across the five foundations, the correlation in the rankings 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.

To investigate whether these results mask heterogeneity at more extreme ideological positions, we present the estimates for the correlation between respondents in each group ($\rho_{g,g'}$) of the seven-category ideological variable (alongside their associated 95% credibility intervals) in figure 5. 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 strong at

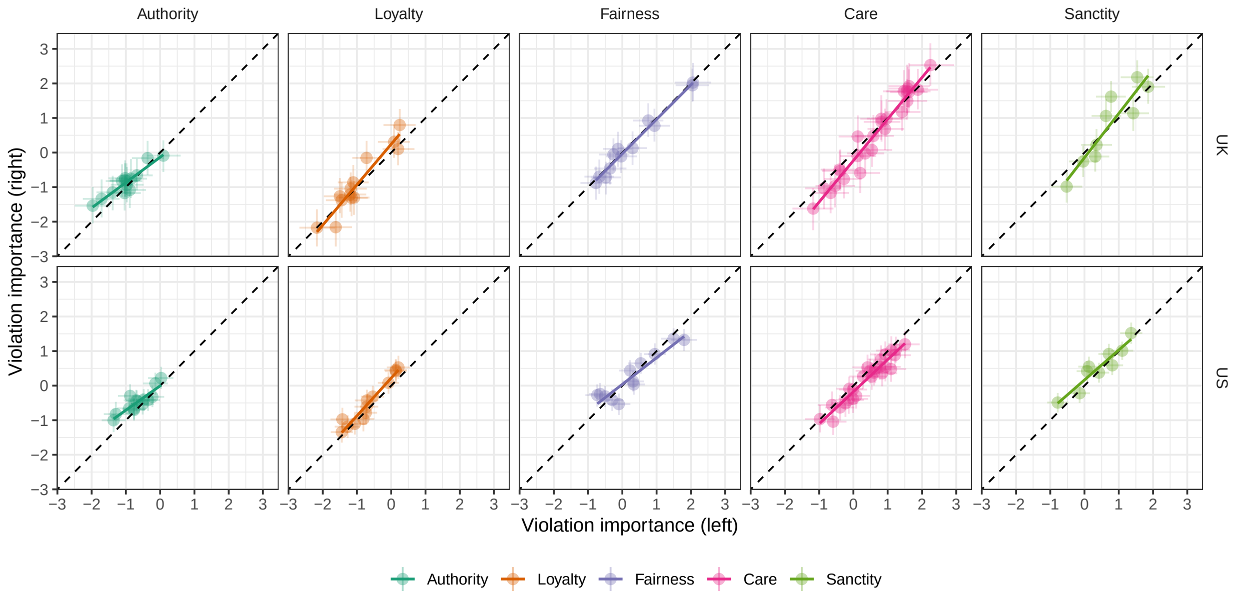


Figure 4: Correlation of violation severity by respondents on the left and right

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]. It is worth emphasizing that the respondents falling into these groups represent small fractions of the population. In the US, only 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 impulses about the relative severity of the violations that we included in our experiment.

We can also re-estimate this correlated severity model to assess whether other voter-level characteristics are associated with different perceptions of violation severity. We find little evidence that this is the case. Most notably, the political voting histories of our respondents seems 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

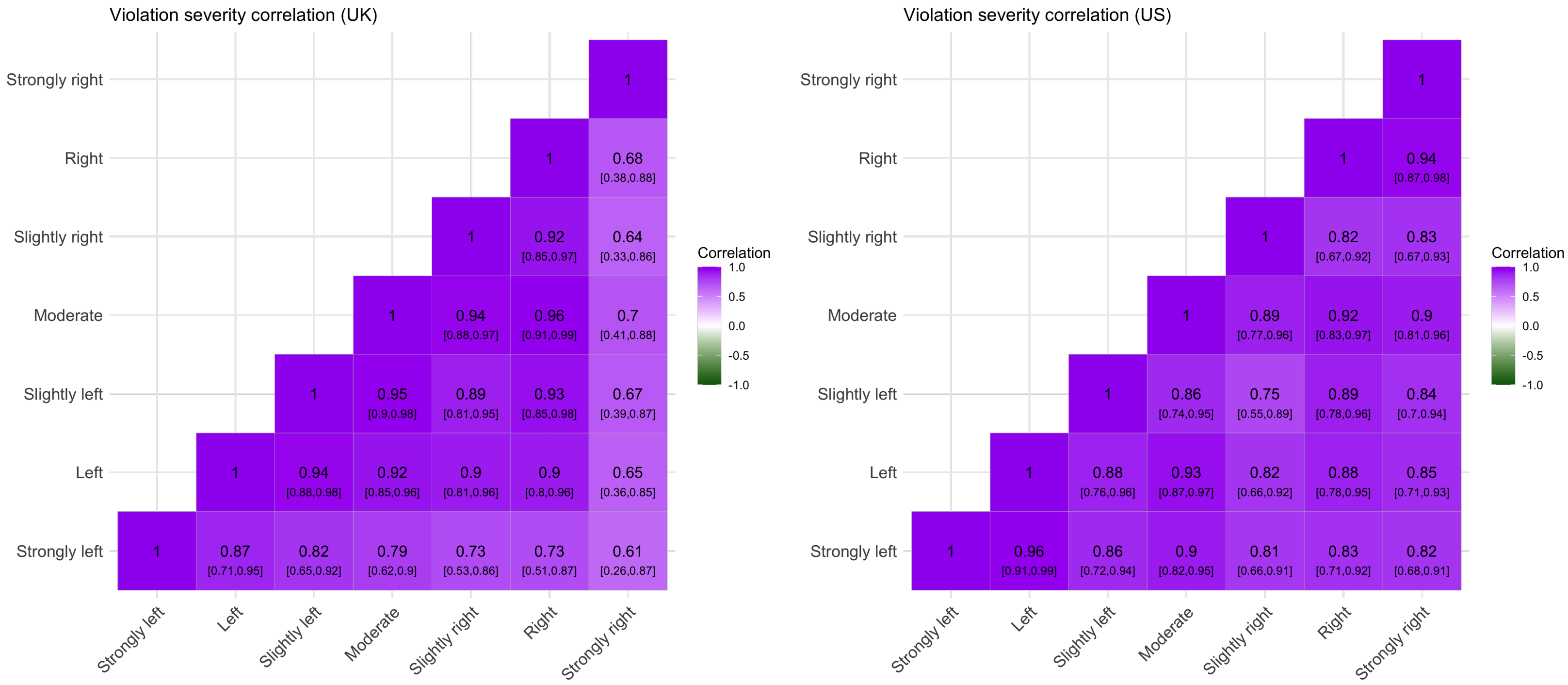


Figure 5: Correlation of violation severity by ideology

Note: The figure shows the estimated correlation in violation severity, across all foundations, for respondents with different self-reported ideological positions.

about which violations are morally worse: the estimated ρ parameter for these groups is 0.95.[18]

In general, whether we use an attitudinal or behavioral measure of political ideology, the fact that we find such high correlations across respondents of different ideological types suggests that 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 intuitions between respondents of different ideological, political or demographic groupings. The evidence here is hard to square with an interpretation that differences in moral judgment 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 reasoning in surveys, when confronted with concrete moral dilemmas they tend to react in strikingly similar ways.

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 widely used approaches for measuring moral priorities. 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 judgment that proponents of MFT believe underpin political disagreement between the left and the right. Empirically, where prior studies find that liberals and conservatives *articulate* different sets of moral priorities, we find

[18] We also find that men and women have highly correlated views on the severity of different violations (UK = 0.99; US = 0.99), and the same is true when comparing those under 35 and those over 65 (UK = 0.93; US = 0.91), and those with post-graduate degree-level education compared with high-school education (UK = 0.96;US = 0.92). We also find that evaluations of violation severity between respondents from the UK and respondents from the US correlate at 0.97. 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 (UK = 0.96;US = 0.96).

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.

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 provide a more optimistic conclusion about the potential for productive conversation across political lines of disagreement. If conservatives and liberals react in largely similar ways to concrete moral questions (as we show here), but express much more variation in their self-assessed moral attitudes (as documented in existing work), then the latter may reflect differences in how people talk about moral questions rather than genuine moral conflict. This suggests that compromise and conciliation across political lines is more tractable than previously suggested.

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 of our vignettes is helpful because it reduces the probability that people will 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 are likely to know what people of their political group tend to think on such issues. In fact, our findings suggest that where prior work finds such political differences, they are unlikely to stem from fundamentally incompatible moral views on the importance of sanctity (or another foundation), but rather are a product of something else. 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 too small to be responsible for the well-documented polarization between ideological groups in advanced democracies.

Appendix 1 – Moral Violation Vignettes

Table 3: Moral violation texts

UK text	US alternative
Authority	
A girl/boy ignoring her/his father's orders by taking the car after her/his curfew.	
A girl/boy repeatedly interrupting her/his teacher as he explains a new concept.	
A girl/boy turning up the TV as her/his father talks about his military service.	
A group of women/men having a long and loud conversation during church.	
A player publicly yelling at her/his football coach during a game.	A player publicly yelling at her/his soccer coach during a game.
A star player ignoring her/his coach's order to come off the pitch during a game.	A star player ignoring her/his coach's order to come to the bench during a game.
A student stating that her/his professor is a fool during an afternoon class.	
A teaching assistant talking back to the teacher she/he is assisting in the classrom.	
A teenage girl/boy coming home late and ignoring her/his parents' strict curfew.	
A woman/man secretly watching sport on her/his mobile phone during church.	A woman/man secretly watching sports on her/his cell phone during church.
A woman/man talking loudly and interrupting the mayor's speech to the public.	
A woman/man turning her/his back and walking away while her/his boss questions her/his work.	
An employee trying to undermine all of her/his boss' ideas in front of others.	
An intern disobeying an order to dress professionally and comb her/his hair.	
Care	
A girl/boy laughing at another student forgetting her/his lines at a school play.	
A girl/boy laughing when she/he realizes her/his friend's dad is the cleaner.	
A girl/boy making fun of her/his brother for getting dumped by his girl/boyfriend.	
A girl/boy placing a drawing pin sticking up on the chair of another student.	A girl/boy placing a thumbtack sticking up on the chair of another student.
A girl/boy saying that another girl/boy is too ugly to be in the class photo.	
A girl/boy setting a series of traps to kill stray cats in her/his neighborhood.	
A girl/boy telling a boy that his older brother is much more attractive than him.	

Table 3: Moral violation texts *(continued)*

UK text	US alternative
A girl/boy telling a woman/man that she/he looks just like her/his overweight bulldog.	
A girl/boy telling her/his classmate that she/he looks like she/he has gained weight.	
A girl/boy throwing rocks at cows that are grazing in the local field.	
A teacher hitting her/his student's hand with a ruler for falling asleep in class.	
A teenage girl/boy chuckling at an amputee she/he passes by.	
A teenage girl/boy openly staring at a disfigured woman/man as she/he walks past.	
A woman/man clearly avoiding sitting next to an obese woman/man on the bus.	
A woman/man commenting out loud about how fat another woman/man looks in her/his jeans.	
A woman/man lashing her/his pony with a whip for breaking loose from its pen.	
A woman/man laughing as she/he passes by a cancer patient with a bald head.	
A woman/man laughing at a disabled co-worker while at an office party.	
A woman/man leaving her/his dog outside in the rain after it dug in the rubbish.	A woman/man leaving her/his dog outside in the rain after it dug in the trash.
A woman/man loudly telling her/his husband that the dinner he cooked tastes awful.	
A woman/man quickly canceling a blind date as soon as she/he sees the man.	
A woman/man spanking her/his child with a spatula for getting bad grades in school.	
A woman/man swerving her/his car in order to intentionally run over a squirrel.	
A woman/man telling a man that his painting looks like it was done by children.	
A woman/man throwing her/his cat across the room for scratching the furniture.	
A zoo trainer jabbing a dolphin to get it to entertain her/his customers.	
Fairness	
A girl/boy skipping to the front of the queue because her/his friend is an employee.	A girl/boy skipping to the front of the line because her/his friend is an employee.
A judge taking on a criminal case although she/he is friends with the defendant.	
A politician using public money to build an extension on her/his home.	A politician using federal tax dollars to build an extension on her/his home.

Table 3: Moral violation texts *(continued)*

UK text	US alternative
A professor giving a bad grade to a student just because she/he dislikes him.	
A referee intentionally making bad decisions that help her/his favoured team win.	A referee intentionally making bad calls that help her/his favoured team win.
A student copying her/his classmate's answer she/heet on an exam.	
A tenant bribing her/his landlord to be the first to get her/his flat repainted.	A tenant bribing her/his landlord to be the first to get her/his apartment repainted.
A woman/man cheating in a card game while playing with a group of strangers.	
A woman/man lying about the number of paid days off she/he has taken from work.	A woman/man lying about the number of vacation days she/he has taken at work.
A woman/man playing football pretending to be seriously fouled by an opposing player.	A woman/man playing soccer pretending to be seriously fouled by an opposing player.
A woman/man taking a shortcut on a running course during a race in order to win.	
An employee lying about how many hours she/he worked during the week.	
Loyalty	
A British film star saying she/he agrees with a foreign dictator's denunciation of the UK.	An American movie star saying she/he agrees with a foreign dictator's denunciation of the US.
A British woman/man telling foreigners that the UK is an evil force in the world.	An American woman/man telling foreigners that the US is an evil force in the world.
A coach celebrating with the opposing team's players who just won the game against her/his team.	
A former politician publicly giving up her/his citizenship to the UK.	A former politician publicly giving up her/his citizenship to the US.
A former UK Army General saying publicly she/he would never buy any UK product.	A former US General saying publicly she/he would never buy any American product.
A local politician saying that the neighboring town is much better than her/his town.	
A teacher publicly saying she/he hopes another school wins a competition.	
A UK swimmer cheering as a Chinese foe beats her/his teammate to win the gold.	A US swimmer cheering as a Chinese foe beats her/his teammate to win the gold.
A woman/man leaving her/his family business to go work for their main competitor.	
A woman/man secretly voting against her/his husband in a local talent competition.	
An employee joking with competitors about how badly her/his company did last year.	
The head girl/boy saying that her/his rival school is a better school.	The class president saying that her/his rival college is a better school.
The UK Ambassador joking in America about how stupid she/he thinks the British are.	The US Ambassador joking in Great Britain about how stupid she/he thinks Americans are.

Table 3: Moral violation texts *(continued)*

UK text	US alternative
Sanctity	
A drunk elderly woman/man offering to have oral sex with anyone in the bar.	
A single woman/man ordering an inflatable sex doll that looks like her/his assistant.	
A small group of religious women/men eating the flesh of their naturally deceased members.	
A woman/man having sex with a frozen chicken before cooking it for dinner.	
A woman/man in a bar offering sex to anyone who buys her/him a drink.	
A woman/man in a bar using her/his phone to watch people having sex with animals.	
A woman/man marrying her/his first cousin in an elaborate wedding.	
A woman/man searching through the rubbish to find men's discarded underwear.	A woman/man searching through the trash to find men's discarded underwear.
An employee at a morgue eating her/his pepperoni pizza off of a dead body.	

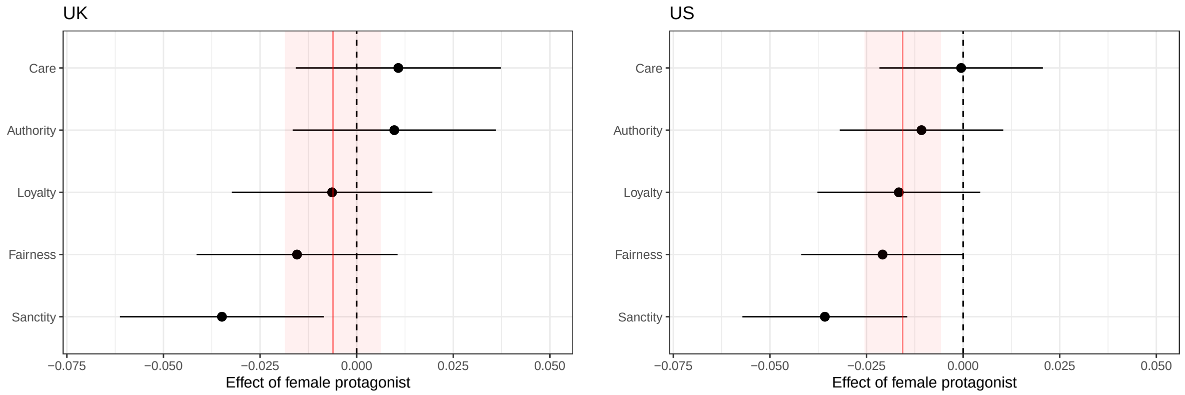


Figure 6: The figure shows the effect of a female protagonist on the probability of a violation being selected as "worse". Negative values indicate that violations describing actions by women are selected as "worse" less often than those describing actions by men.

Appendix 2 – Gender-based treatment effects

We include two versions of each violation described in Clifford et al. (2015) in our experiment: one in which the person committing the moral violation is a man, one where it is a woman. In this section we analyse whether the gender of the protagonist affects perceptions of the severity of these violations.

In figure 6 we plot the results from two regressions – one using the UK sample and one using the US sample – in which we stack the data such that we have two observations for each pairwise comparison, and an outcome variable that is coded 1 if a given violation was selected as "worse" in the comparison, 0 if the other violation was selected as worse, and 0.5 if they were considered to be "about the same". We then regress this outcome on a binary indicator which measures whether the protagonist in the violation was a woman, rather than a man. We also interact this indicator with dummy variables which capture the foundation to which a given violation applies, allowing us to assess whether these gender effects vary according to violation type.

Figure 6 depicts the marginal effect of the female protagonist treatment for each of the five foundations, alongside 95% confidence intervals. The red line represents the average (across foundation) effect of the female protagonist treatment, with 95% confidence bands shaded also in red. The figure shows that, on average, respondents are somewhat less likely to select violations perpetrated by women as “worse”, but this effect is only significantly distinguishable from zero in the US data, and the effect is small in magnitude.

Decomposing these effects by foundation reveals that it is especially with regard to the sanctity foundation that violations by women that are perceived to be less severe than violations perpetrated by men. In both the UK and US data, the treatment effect of a female protagonist is negative and significantly different from zero. This suggests that while the sanctity violations are considered by our respondents to be among the most severe moral wrongs we include in our experiment (as evidenced by the findings in the main text), such transgressions are considered somewhat less objectionable when committed by women than by men.

Finally, we also investigated whether these effects vary according to the gender of the *respondents*. We might expect, for instance, that female respondents are likely to select moral violations by female protagonists as “worse” less often than male respondents. However, in both the UK and US data we find that respondent-gender-by-violator-gender interactions are insignificant. This suggests that perceptions of violation severity are not strongly conditioned by gender, either of the people described in the vignettes, or of the respondents making the comparisons.

References

Abramowitz, Alan I and Kyle L Saunders. 2008. "Is polarization a myth?" *The Journal of Politics* 70(2):542–555.

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.

Ciuk, David J and William G Jacoby. 2015. "Checking for systematic value preferences using the method of triads." *Political Psychology* 36(6):709–728.

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.

Day, Martin V, Susan T Fiske, Emily L Downing and Thomas E Trail. 2014. "Shifting liberal and conservative attitudes using moral foundations theory." *Personality and Social Psychology Bulletin* 40(12):1559–1573.

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

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.

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

Graham, Jesse. 2010. Ideology and Automatic Moral Reactions PhD thesis University of Virginia.

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.

Hainmueller, Jens, Dominik Hangartner and Teppei Yamamoto. 2015. "Validating vignette and conjoint survey experiments against real-world behavior." *Proceedings of the National Academy of Sciences* 112(8):2395–2400.

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.

Heekeren, Hauke R, Isabell Wartenburger, Helge Schmidt, Kristin Prehn, Hans-Peter Schwintowski and Arno Villringer. 2005. "Influence of bodily harm on neural correlates of semantic and moral decision-making." *Neuroimage* 24(3):887–897.

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* .

Hobolt, Sara B, Thomas J Leeper and James Tilley. 2021. "Divided by the vote: Affective polarization in the wake of the Brexit referendum." *British Journal of Political Science* 51(4):1476–1493.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science* 22:129–146.

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.".

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.

Kohlberg, Lawrence, Charles Levine and Alexandra Hewer. 1983. "Moral stages: A current formulation and a response to critics.".

Kolberg, Lawrence. 1984. *The Psychology of Moral Development*. San Francisco: Harper and Row.

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.

Kunda, Ziva. 1990. "The case for motivated reasoning." *Psychological bulletin* 108(3):480.

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.

Sagi, Eyal and Morteza Dehghani. 2014. "Measuring moral rhetoric in text." *Social science computer review* 32(2):132–144.

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.

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.

Turiel, Elliot. 1983. *The development of social knowledge: Morality and convention*. Cambridge University Press.

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