Human Rights Violations and Public Support for Sanctions*

Barış Arı† Burak Sonmez‡

December 21, 2021

ABSTRACT

Public pressure to take punitive action against human rights violators is often a driving force behind international sanctions. Yet, we know little about how public support is shaped by varying types of abuse, the costs and effectiveness of sanctions, and the differential harm they inflict upon the target population and leadership. Our study specifically addresses this gap by unpicking contextual factors that jointly sway the perception of morality and the cost-benefit calculus. Findings from our paired conjoint experiment suggest that different categories of human rights abuses have varying degrees of perceived salience to merit international sanctions. Individuals also prefer sheltering the target population while punishing the leadership, but collective punishment becomes less unacceptable if majority of the target population supports the human rights infringements. The desire to do something against the perpetrators amplifies the appeal of punishing the leadership but assuages the moral concerns of harming the population.

Keywords: Human Rights; Sanctions; Experiment; Public Opinion

JEL-Codes: F51, C99, D91

---

*The authors contributed equally to the study. This work was supported by the Economic and Social Research Council [grant number ES/T006013/1]; and Durham University. Competing interests: Authors declare no competing interests; Data and materials availability: All data, code, and materials used in the analysis will be made available via Open Science Framework (OSF) repository. The authors thank Patrick Kuhn, Rebecca Cordell, Patrick Leslie for useful comments and suggestions.

† Barış Arı (baris.ari@uea.ac.uk – ORCID: **0000-0003-1300-8880**) is Lecturer in Political Science at School of Politics, Philosophy, Language and Communication Studies, University of East Anglia, Norwich, NR4 7TJ, UK

‡ Burak Sonmez (b.sonmez@ucl.ac.uk – ORCID: **0000-0002-2335-566X**) is Lecturer in Quantitative Social Science in UCL Social Research Institute at University College London, London WC1H 0AA, UK

Imposing economic sanctions on human rights abusing countries is a common foreign policy response in democracies. Yet, the use of sanctions is beyond an instrumental tool to advance foreign policy goals and promote human rights abroad. Governments often use sanctions to appease the demands of domestic constituencies (Whang 2011; McLean and Whang 2014; Kustra 2021). For example, bowing to the demands of his Christian support base, President George Bush imposed sanctions on Sudan as a response to gross human rights violations in Darfur (Goldenberg 2007). Similarly, the Obama administration imposed sanctions on Uganda in 2014 by declaring that the introduction of anti-gay laws was "counter to universal human rights" (BBC 2014). As policymakers in democracies are receptive to demands of the public, campaigners aiming to instigate action against human rights violators face the task of mustering public support. This link renders studying micro-foundations of citizen support particularly important in understanding international sanctions because incentives to promote human rights abroad are also rooted in political considerations at home.

In this study, we investigate the factors that influence citizen willingness to impose sanctions on human rights violating countries. Our starting point is that an individual reflects on instrumental and moral dimensions together as a whole when forming her opinion. Studies grounded on the cost-benefit framework underscore the costs of sanctions on the sender country and their effectiveness to induce compliance of the receiver as the two primary dimensions that the public considers (Heinrich, Kobayashi, and Peterson 2017; Putnam and Shapiro 2017). A competing approach draws attention to the dimension of morality to argue that normative considerations, as opposed to instrumental concerns, can sway the public opinion to pursue costly foreign policy tools without significant material benefits (Kreps and Maxey 2018). Regardless of their

effectiveness at securing compliance, sanctions may have a costly but expressive purpose of reinforcing morality (Galtung 1967, 412).

Building on this debate, we identify contextual factors that influence citizen opinion. Our key argument is that there is no simple separation between instrumental and moral considerations. When forming their opinion, an individual makes a multidimensional trade-off by reflecting on several contextual factors that jointly affect the perception of morality as well as the cost-benefit analysis (Heinrich and Kobayashi 2020). Particularly, we identify the cost of sanctions on the receiver in terms of type and volume (i.e., who is hurting and to what extent) as an influential –but insufficiently investigated— factor.

We contribute to the empirical study of public opinion formation by recognizing the difference between collective and targeted sanctions. Although this difference is central both in public debate and foundational theory (Galtung 1967; Weiss 1999; Drezner 2011), studies investigating the micro-foundations of support for foreign policy responses to human rights abuses have not considered the cost of intervention that falls on the public of the targeted country as a moral consideration. As Kirshner (1997, 33) argues, a simple distinction between the sender and the target is insufficient: "instead of considering how [economic] sanctions hurt the target state", research on micro-foundations should focus on "how groups within the target are affected differentially, and how these consequences change with the form of statecraft chosen." We explicitly incorporate such an essential dimension into our experimental design and further propose that citizens use the information on the type of abuse and the political context within which human rights violations occur when they attribute individual- or collective-level accountability.

Our framework sheds further light on the mixed findings in the empirical literature regarding public support for foreign policy instruments. On the one hand, an emerging line of research challenged the conventional wisdom that public opinion is driven by normative concerns by arguing that individuals are more self-serving and goal-oriented than previously assumed (Christiansen, Heinrich, and Peterson 2016; Heinrich, Kobayashi, and Long 2018; Heinrich and Kobayashi 2020). On the other hand, several recent studies also found that individuals tend to prioritize the humanitarian over the instrumental when asked about military interventions (Kreps and Maxey 2018; Tomz and Weeks 2020). Studying interactions among prominent issue dimensions while taking contextual factors into account is a crucial step forward for addressing this puzzle.

Following this line of thought, we carried out a pre-registered conjoint experiment to simultaneously test the impact of previously omitted factors on citizen preferences toward economic sanctions to promote human rights. Holding the volatile political environment constant, we reveal a number of key factors –and interactions among them– in understanding the *preference-based third-party punishment*. More specifically, we consider the multi-dimensionality of morality, the costs of promoting human rights, signaling the likelihood of success of sanctions, conditional cooperation, and the juxtaposition of norm enforcement and norm diffusion.

Our most prominent findings can be summarized under three points. First, we find that individuals do differentiate the target population from the leadership and show a clear preference toward punishing the latter more severely while sheltering the former from the harms of economic sanctions. However, this aversion to harming the target population is dependent on a number of factors. Most notably, when only a small minority of the target population supports the human rights abuses of their transgressing government, respondents are even more averse to inflicting harm to the target

population, but their willingness to differentiate the target population decrease as the proportion of locals supporting the human rights infringing policies increase. As expected, we also find that respondents are sensitive to incurring costs when imposing sanctions. Contrary to our expectations, however, their unwillingness to incur higher costs remains remarkably consistent irrespective of the harm that falls on the target population. Individuals do not become more magnanimous and less averse to incurring costs when the harm of sanctions on the target population decrease, reflecting the saliency of sanction costs and the limitations of incentives to harbor the target population.

Second, different types of human rights abuses have varying degrees of perceived salience to merit international sanctions. Respondents perceive some type of human rights violations (e.g., torture and ill-treatment by state authorities) to be more worthy of punishment than others (e.g., women's reproductive rights and equal marriage rights). Six distinct types of violations in our experimental design are ranked in three clusters based on their perceived importance to merit sanctions. Moreover, we find that the type of violation and local support for it interact. For example, individuals become particularly more willing to punish restricting language and religious practices of minorities when the target population overwhelmingly supports such transgressions.

Finally, our results present support for the expressive function of sanctions in reinforcing morality. Contrary to our expectations, the ineffectiveness of prior sanctions does not dissuade individuals to issue new rounds of sanctions. Similarly, respondents disregard the effectiveness of prior sanctions when evaluating sanction costs falling on the target population and leadership. Individuals are consistent in their enthusiasm to punish the leadership and with their disapproval of harming the population, irrespective of the prior compliance. Although respondents disregard the effectiveness of prior

sanctions, they show a preference toward maiden sanctions (i.e., imposing sanctions on a country that had not received sanctions before). This desire to do something against the offenders may even amplify the appeal of punishing the leadership, and more surprisingly, assuage the moral concerns of harming the target population.

The rest of this article is organized as follows. First, we provide a brief review of studies investigating the micro-foundations of public preferences toward costly foreign policy instruments to pursue humanitarian goals. We particularly focus on studies with experimental design as our main interest is citizen decision-making. We continue with our theoretical framework in which we derive a series of hypotheses. After explaining our experimental design, we present our most noteworthy results. In the concluding remarks, we highlight the governmental and non-governmental policy implications of our study and explore areas of further research.

Public Opinion and Foreign Policy Responses to Human Rights Violation

Public opinion matters for foreign policy decisions, at least in democracies. Leaders often seek domestic support for punitive actions against a foreign country by appealing to the humanitarian values of the public. Such rhetoric can be even used to justify instrumental actions driven by material and strategic goals. Conversely, otherwise uninterested policymakers may find themselves under pressure from the public to do something against international human rights violators. An established research tradition showed that human rights organizations, the media, voters, and special interest groups are all influential on how leaders craft and implement foreign policy instruments (Whang 2011; Murdie and Peksen 2013; McLean and Whang 2014; Peksen, Peterson, and Drury 2014). This strong link between domestic public opinion and foreign policy renders the study of micro-foundations of punitive foreign policy instruments particularly important.

Despite its importance, only small but growing literature investigated how contextual factors and key issue dimensions affect individual opinion formation through a causal framework. In the specific context of economic sanctions, Putnam and Shapiro (2017) and Heinrich et al. (2017) are the only studies using an experimental design. In line with the academic debate on economic sanctions, both studies consider costs on the sender as a central dimension that individuals would take into account when forming their opinion. As expected, higher costs of sanctions on the sender decrease respondent support. Heinrich et al. (2017) further investigate both the short- and long-term effectiveness of sanctions and find a positive relationship. Neither research, however, varies the type of human rights abuse in their experimental design.[1] Considering that individuals tend to value some human rights more than the others depending on their beliefs and political preferences (McFarland and Mathews 2005), the type of human rights violation is a crucial dimension. Similarly, whether the sanctions inflict harm to the leaders of the violating country or the general populace is central for the micro-foundations of economic sanctions (Kirshner 1997). Yet, the differential cost of sanctions on the target has not been investigated in an experimental design. For example, Heinrich et al. (2017) consider the cost of sanctions on the target as a measure of the severity of punishment, but their experimental design does not differentiate how these costs are distributed within the target.

One strand of the literature focus on the ways in which American and British citizens operate moral qualms about waging war against a country that violated human rights (Kreps and Maxey 2018; Tomz and Weeks 2020). Since individuals tend to have strong moral reactions against human sufferings and severe wrongdoings, the public is

[1] Only Putnam and Shapiro (2017) gauge the severity of human rights violation by worsening the conditions of forced labor.

more likely to justify retribution against human rights violators (Wheeler 2000; Stein 2015). Following the expansion of human rights through universal declarations, international humanitarian norms enable and encourage citizens and their states to become morally responsible against mistreated strangers (Finnemore 1996). Nonetheless, the experimental literature often presents human rights violations abroad through wartime-like scenarios by setting the contextual factors aside and by ignoring the nuances in morality, which are linked with diverse types of violations. This generates limitations for studying individual opinion formation more broadly, especially considering that a theory of norms cannot leave the specific social context out of consideration (Granovetter, 1985).

Recent studies investigating micro-foundations isolate some contextual factors affecting public support for action against human rights violators, but they do not disentangle how citizen preferences change when the characteristics of the targeted country, the type of human rights violation, and differential costs of enforcing sanctions vary. Similarly, a strand of the literature sheds light on how norm compliance with international human rights is influenced by the existence of international obligations (Chilton 2014; 2015; Chilton and Versteeg 2016), but the extent to which public opinion is characterized by a specific moral duty to promote global human rights, irrespective of any external legal enforcement mechanism, remains out of the scope.

In this respect, further experimental research is needed not only to expand our understanding of contextual factors and interactions among them, but also to validate earlier findings. The latter is needed because single random assignment in experimental studies brings methodological caveats, such as obstacles to isolate the mediating causal pathways of potential intervening variables. Further experimental research is needed also because comparing different types of human rights abuses while holding all other

attributes constant to infer humanitarian motives is also not empirically viable using observational data (Druckman 2011).

Theoretical Framework and Hypotheses

Next, we turn to our theoretical framework which helps us to identify crucial contextual factors and informs our experimental design. We propose that varying the severity, kind, and target of human rights abuses is necessary to uncover invaluable insights in understanding different aspects of morality (Conrad, Hill, and Moore 2018).

First of all, citizens prioritize or are more supportive of the protection of certain human rights (e.g., conventions against cruel treatment) due to the perceived severity associated with the type of abuse and politically motivated reasoning (Whitmeyer 2002). The context within which violations take place is likely to induce different responses. For example, individuals may consider human rights that are universally endorsed more worthy of protection, such as freedom from torture and slavery. In this vein, an established research tradition also shows that attitudes toward human rights protection vary among individuals in line with their ideological and socioeconomic features (McFarland and Mathews 2005). Based on this conjecture, we argue that the type of abuse matters and formulate the following hypotheses.

Hypothesis 1.1: *Violations regarding torture and cruel treatment by state authorities will draw higher support for sanctions compared to the type of violations unevenly endorsed by every spectrum of society, such as restriction of minority language and religious practices, women's reproductive rights, and equal marriage rights.*

Hypothesis 1.2: *Violations regarding forced child labor will draw higher support for sanctions compared to restriction of minority language and religious practices, women's reproductive rights, and equal marriage rights.*

Hypothesis 1.3*: Minority language and religious rights, women's reproductive rights, and equal marriage rights will be preferred as more morally worthy to protect by those whose partisanship or political leaning is more left-wing.*

Hypothesis 1.4*: Women's reproductive rights will be perceived as more morally imperative to protect by those whose sex is female.*

Cost of sanctions and their expected effectiveness

Imposing sanctions to protect human rights entails costs for the sender country and its citizens. Relatedly, the literature expects that public support for any foreign policy instrument is driven by the extent of potential cost as well as the predicted success of the action (Gartner 2008; Heinrich, Kobayashi, and Peterson 2017; Heinrich, Kobayashi, and Long 2018). Therefore, to better comprehend the account of bounded rationality in understanding social preferences, we need to disentangle how individuals make (ir)rational calculations about their choices and resulting trade-offs.

Hypothesis 2.1*: If the receiver country's previous compliance rate with sanctions is low, the public will be less likely to promote human rights through sanctions.*

Hypothesis 2.2*: The higher the costs of sanctions to the sender country, the lower the support for economic sanctions.*

When discussing the trade-offs from implementing an instrument to influence the human rights record of a foreign country, the literature has focused on the costs that fall on behalf of the home country and the likely benefits drawn from such actions in terms of influencing the behavior of the violator country. However, such an approach ignores the costs that fall on behalf of the public (as opposed to policymakers) of the target country. For example, Kreps & Maxey (2018) consider the cost of intervention on the intervenor, but not the civilian costs due to the humanitarian intervention itself. It is reasonable to expect that any humanitarian intervention, however well-designed and executed, would inadvertently generate civilian casualties. Similarly, public opinion research on economic sanctions excludes the cost of economic sanctions on the civilian population as a factor that influences the moral considerations of respondents. The costs that fall on behalf of the violator country are understood only as a factor reflecting the

severity of punishment and the expected effectiveness of sanctions (Heinrich, Kobayashi, and Peterson 2017; Putnam and Shapiro 2017).

We argue that the humanitarian cost on the target country is an essential component of the moral-instrumental trade-off. Indeed, this is a focal point of discussion in theory and public debate regarding economic sanctions (Galtung 1967; Malloy 1995; Weiss 1999). Normative considerations regarding the collateral damage that sanctions inflict upon the civilian population have influenced how countries design and implement sanctions (Drezner 2015). Most notably, human rights organizations often emphasize the widespread suffering among civilian populations and question the ethics of imposing collective sanctions (McFarland and Mathews 2005; Wike and Schumacher 2020). As human rights organizations are influential in shaping public debate and raising citizen awareness that in turn affect the use of sanctions (Whang 2011; Murdie and Peksen 2013; Peksen, Peterson, and Drury 2014), we consider this debate central for studying public opinion. We therefore explicitly incorporate it into our theoretical framework.

We use the distinction between general and targeted sanctions to further unpack the costs on behalf of the receiver. In general, targeted sanctions are designed to generate costs on specific individuals while sparing civilians. Our key argument is that there are two facets of sanction costs: those that fall on the sender, which we call incurred costs, and those that fall on the received, which we call exported costs. We argue that citizens of the sender country are more willing to incur costs of imposing sanctions as long as exported costs fall largely upon the leadership, but not on the general populace. We formulate the following hypotheses.

Hypothesis 3.1*: The higher the costs of sanctions on the general populace of the target, the lower the support for sanctions.*

Hypothesis 3.2: *The higher the costs of sanctions on the leaders of the violator country, the higher the support for sanctions.*

Hypothesis 3.3: *Individuals will be more willing to accept incurred costs if the exported cost on the target population is low.*

The effectiveness of sanctions to influence violator behavior is an important consideration both in academic studies and the public debate. Its critics formulate sanctions as an expensive but ineffective foreign policy tool, which can even bring detrimental results in terms of democracy promotion and human rights protection (Morgan and Schwebach 1997; Wood 2008; Peksen 2009). Others highlight that regardless of their effectiveness, sanctions play an expressive role, addressing the moral responsibility to do something as a response to norm violations (Galtung 1967; Whang 2011). An interaction between the cost and effectiveness is expected to result in low-cost symbolic sanctions that play an expressive role when influencing the behavior of the target is unlikely (McLean and Whang 2014). Addressing this debate, public opinion research also locates the effectiveness as a central factor in the cost-benefit calculation, with an expectation that support for costly sanctions should be low if the target is unlikely to change course, but if the costs incurred are small, sanctions may play an expressive function regardless of their effectiveness (Heinrich, Kobayashi, and Peterson 2017).

Explicitly recognizing that sanctions may harm civilians of the target country without achieving policy change help us further unpack the complex cost-benefit analysis that individuals face. We argue that the difference between general and targeted sanctions is again central in this trade-off, as individuals are more willing to incur costs to impose sanctions to those who they attribute responsibility for human rights violations, even if such sanctions are unlikely to cause behavior change. However, sanctions that affect the general public of the target country without having an impact on

its leaders are unlikely to play an expressive role and muster citizen support. Based on this reasoning, we formulate the following hypotheses.

Hypothesis 4.1: *When the target is unlikely to stop human rights violations, individuals will support sanctions if the exported cost on the leadership is high.*

Hypothesis 4.2: *When the exported cost on the target population is high, individuals will support sanctions if the target is likely to stop human rights violations.*

Norm Diffusion

Establishing shared moral judgments is culturally and politically bounded, thus whether a specific human right is homogenously endorsed or not is a focal point of interest to investigate the causal pathways for supporting human rights protection abroad. Most notably, public attitudes toward supporting a distinct type of human rights significantly vary across countries and people who affiliate themselves with different political ideologies (McFarland and Mathews 2005; Wike and Schumacher 2020). For example, a significant majority of the public in both U.S. and UK endorse the importance of equal rights to practicing religion freely, whereas only 18% of Japanese citizens do support such a right. Drawing on this challenge, third-party punishment for international human rights violations does not simply capture the social preference to enforce cooperation between the norm violator and the protector but may also forcibly alter the moral domain of 'others'. In this regard, norm entrepreneurs with principled ideas, such as human rights advocates, are central in initiating and extending new behavior through the diffusion of international norms (Finnemore and Sikkink 1998; Hyde 2011).

In addition to the heterogeneity in attitudes toward human rights amongst the public of the sender country, a similar heterogeneity exists within the target as well as between states. For example, the conventions of equal treatment for detainees or labor rights are recognized by the majority of UN members whereas the majority of states do

not recognize the full scope of LGBTI rights (Human Rights Watch 2019). This raises the question to what extent democratic citizens are willing to extend their third-party punishment to the context of violations where the majority of sanction-receiving country citizens morally justify the addressed human rights abuse. Since the relevant literature has yet to unveil such questions to our best knowledge, we prefer remaining exploratory in understanding this question. Nonetheless, given the possibility of the within-design interaction with the type of human rights in our experimental design, we also want to test the hypotheses that:

Hypothesis 5.1: *The public is more likely to favor sanctioning when the majority of target country citizens are not against the certain human rights abuses that are more universally accepted, namely torture and ill-treatment by state authorities, forced child labor, and violation of freedom of expression.*

Hypothesis 5.2*: The public is less likely to favor sanctioning when the majority of the target country citizens are not against the certain human rights abuses that are also contentious at home, namely reproductive rights and equal marriage rights.*

When making decisions regarding sanctions that would cause collective harm in the target country, democratic citizens may justify their preferences depending on whether the considerable number of citizens in the target endorse the abusing policy. The third-party punishment for groups is likely to be based on the perceived behavior of the majority rather than a minority of policymakers in such collective actions. Following this argument, we revisit the discussion based on simulations that if people anticipate lower levels of compliance, sanctioning on the group rather than individuals may be seen as the most cost-effective means (Whitmeyer 2002). Although the public in democratic countries should have an aversion toward harming civilians through sanctions, this disinclination is likely to be conditional on the target population's perceived level of support toward the human rights abusing policies. We, therefore, hypothesize that:

Hypothesis 5.3*: Public will become more likely to accept imposing costs on the target population as the proportion of the population supporting the abuse increases.*

Conditional Co-operation

Drawing on the conditional cooperation theory in behavioral social sciences (Cheung 2014), we assume that citizens condition their public preferences on the expected behavior of other states. Note that promoting global human rights involves a certain collective action dilemma where acting in response to human rights violations is costly, thus generating the expectation of free riding. More precisely, if third-party countries are expected to contribute to protecting human rights through complying with sanctions, which signals that free-riding is not a severe challenge, then it is reasonable to behave similarly and be willing to sacrifice in the hope of a better payoff (Cheung 2014). Our hypothesis is therefore as follows:

Hypothesis 6.1: *The public is more likely to sanction when the proportion of conditional cooperation is relatively higher.*

Experimental Design

Our empirical strategy is based on a Web-based paired conjoint experiment with forced-choice design to examine the heterogeneity of third-party punishment regarding the violations of global human rights (see Appendix A1). Unlike a binary treatment in traditional survey experiments, we simultaneously manipulate seven different attributes of human rights violations (see Table 1): the type of human rights abuse; the costs of sanctions on the sender; the harm inflicted on civilians and the leadership of the target; the previous compliance behavior toward sanctions; the public endorsement of human rights abuse; and behaviors of third-party states. We thereby offer a more comprehensive understanding of citizens' preferences for sanctions.

Table 1. Imposing Economic Sanctions in Conjoint Profiles

Attributes	*Features*
Type of Human Rights Abuse	{Torture and ill-treatment in detention; Restricting freedom of expression by censoring publications and the internet; Forced child labor; Restricting language and religious practices of minorities; Restricting abortion rights; Restricting equal marriage rights}
Cost of the Sanction on the U.S./UK Households	{High food price Inflation; Low food price inflation}
Cost of the Sanction on the Leadership of Target	{Major economic harm; Minor economic harm}
Cost of the Sanction on the General Population of Target	{Major economic harm; Minor economic harm}
Whether the Country Previously Stopped Abuse After Earlier Sanctions	{Stopped; Didn't stop; Never sanctioned before}
The Percentage of Citizens in the Targeted Country Supporting the Abusive Policy	{less than 10%, around 50%, more than 90%}
The Number of Countries Supporting the Sanction	{0 out of 192; 20 out of 192; 80 out of 192; 170 out of 192}

The conjoint design allows respondents to choose or rate two or more hypothetical choices that have multiple attributes with the objective of estimating the influence of each characteristic on respondent's preference intensity (Green, Krieger, and Wind 2001; Bechtel and Scheve 2013; Hainmueller, Hopkins, and Yamamoto 2014; Hansen, Olsen, and Bech 2015). Our design accordingly presents respondents with two different hypothetical human rights abusing country profiles, which rotate through a random set of attributes. With respect to the within-subjects design, each respondent is given five pairs of country profiles to evaluate during the experiment.

Such a paired conjoint design brings several advantages to mimic real-life decision-making. Most notably, examining prominent issue dimensions within their respective contexts, as well as interactions amongst them, enable unpacking the complex

trade-offs that individuals face. Moreover, human rights abuses are unfortunately ubiquitous in any given moment in recent history, as many countries have violated the rights of their citizens. It is reasonable to assume that not each and every abuser country can be targeted through sanctions or humanitarian interventions. Human rights campaigners often try to influence public opinion and pressure governments to act against abuser countries, but given the ubiquitous nature of violations, activists focusing on a specific issue or country may find themselves in an unwanted competition for public attention with other campaigners with a different country or issue in focus. Indeed, comparisons between violator countries emerge frequently in public discussions.[2] A government may even use sanctions on one violator country to divert the attention away from another one.[3] Therefore, we contend that a conjoint design in the context of economic sanctions is advantageous to optimize the capacity to decompose the effects of multidimensional traits on making decisions on imposing sanctions.

The experiment was designed in Qualtrics and performed in July 2020 after being pre-registered at https://osf.io/npkbg and granted ethical approval from Durham University (SGIA-2020-06-19T10:07:14-jx85). Because of the *masking-satisficing* trade-off in conjoint experiments (Bansak et al. 2021), we conducted a pilot study and tested the number of attributes, instructions, their theoretical suitability, the desired level of

[2] One example is human rights activist Harry Wu's comparison of China with Myanmar during a congressional hearing. In response to potential harm of sanctions on the Chinese population, Wu cited sanctions on Myanmar to argue that such concerns were not present when punishing another human rights abusing country (U.S. Congress 1998).

[3] For example, United Kingdom spearheaded international initiatives to impose sanctions on Rhodesia largely to control the agenda and insulate South Africa: "having displayed their outraged morality by calling for sanctions, [UK's] main concern was to steer the Security Council away from any action that would lead to an economic confrontation with South Africa, Britain's fourth largest customer and main supplier of gold" (Time 1966).

realism, and relative strength of the attributes in conjoint profiles before fielding the study.

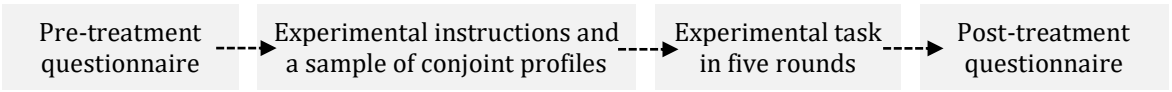


Figure 1. *Overview of the Experimental Design*

Figure 1 summarizes our experimental design. The experimental setup starts with a pre-treatment questionnaire measuring respondents' prior human rights attitudes and commitment. To make sure that respondents comprehend the experimental task, they are evaluated with a question following instructions. Next, the respondents are given two human rights abusing country profiles to evaluate in five rounds. In this evaluation, our main outcome variable measures respondent's preference over imposing economic sanctions on a human rights violating country. More specifically, we ask individuals the following question: "Which of these countries would you like your government to impose economic sanctions the most?". In the final stage, they take a post-treatment questionnaire to provide their demographic information. All questions and question blocks were randomly ordered to avoid spillover effects. We also fully randomized the features of attributes for each country profile each round and the order of attributes across respondents to avoid primacy effects. The experiment was carried out through the participant pool of Prolific Academic, which provided us with a high-quality online opt-in representative sample (see of both the UK (N= 1009) and the U.S. (N=992) based on age, sex, and ethnicity (Peer et al. 2017).

Our experimental design is well-powered for both samples (see Appendix A1.3) and suited for measuring the effect of any sanction characteristic on respondent preferences. It allows us to non-parametrically estimate the effects of different attributes on the support for sanctions, as well as to compare the intensity of support across

different features. In analyzing these effects, we follow the same identification strategy suggested by (Hainmueller, Hopkins, and Yamamoto 2014). We estimate the average marginal component-specific effect (AMCE) as an alternative quantity of interest that gives us the average effect of a change in a country profile on the probability of imposing economic sanctions on the human rights abusing country. Using this estimand, we propose to look at the effect of an individual treatment component. In other words, we are interested in how different values of the l th attribute of profile j influence the probability that the profile is chosen. The effect of attribute l, however, may differ depending on the values of the other attributes. For example, we can be interested in whether respondents tend to choose a certain type of human rights abuse over the other to impose sanctions. We focus on how these treatment effects vary across different characteristics and traits of the research subjects.

Even though AMCEs allow us to disentangle the causal effect of each feature in conjoint profiles, we also report marginal means to describe the level of sanction favorability for all feature levels without being interpreted relative to the baseline categories. Especially for analyzing sub-group preferences, Leeper et al. (2020) demonstrate that conditional AMCEs can be substantially misleading when interpreting the degree of favoring or disfavoring between subgroups since interactions are sensitive to the baseline category used in regression analysis. The effect of type of human rights abuse, however, might also differ depending on whether the public of sanction receiving country approves the violation of that human rights. Therefore, we also analyze within-design interactions and differences in marginal means. In all our analyses, standard errors are clustered by the respondent to avoid biased estimates of the variance, because the respondents are given two country profiles to evaluate in five rounds.

Results

Figure 2 (left panel) reports the estimated AMCEs based on the entire sample along with 95% confidence intervals that show the effect of change in attributes of country profiles on the probability of imposing economic sanctions. Figure 2 (right panel) also reports the marginal means representing the favorability of sanctioning a certain human rights abusing country profile (i.e., the mean outcome across all appearances of a particular feature of human rights abusing country profiles, averaging across all other features). Note that marginal means have a direct interpretation as probabilities: values above 0.5 indicate that the feature increases sanction's favorability whereas values below 0.5 indicate that a decrease in sanction's favorability. In Appendix A2.1, we also present our main results in a table format.

Starting with the cost-benefit analysis of imposing economic sanctions, we find that respondents disfavor sanctions that result in incurring higher costs. Respondents prefer low food price inflation by 4 percentage points ($p < 0.001$), relative to high price inflation. This result confirms Hypothesis 2.2. On the other hand, we find no convincing evidence for Hypothesis 2.1 that if the sanction receiving country's previous compliance with sanctions is low, people are less likely to impose sanctions. We find no meaningful difference between "did not stop" and "stopped" categories albeit the direction of the relationship is contrary to our expectations. The results also indicate that compared to sanctioning a repeat offender that had previously changed course due to the effectiveness of prior sanctions, respondents favor imposing sanctions on a country that had not received any previous sanctions by 3 percentage points ($p < 0.001$). Indeed, the right panel of Figure 2 shows that maiden sanctions is the most preferred category, suggesting that the expressive role of sanctions and the desire to do something against human rights offenders irrespective of the outcome may well be the driving force behind our results.

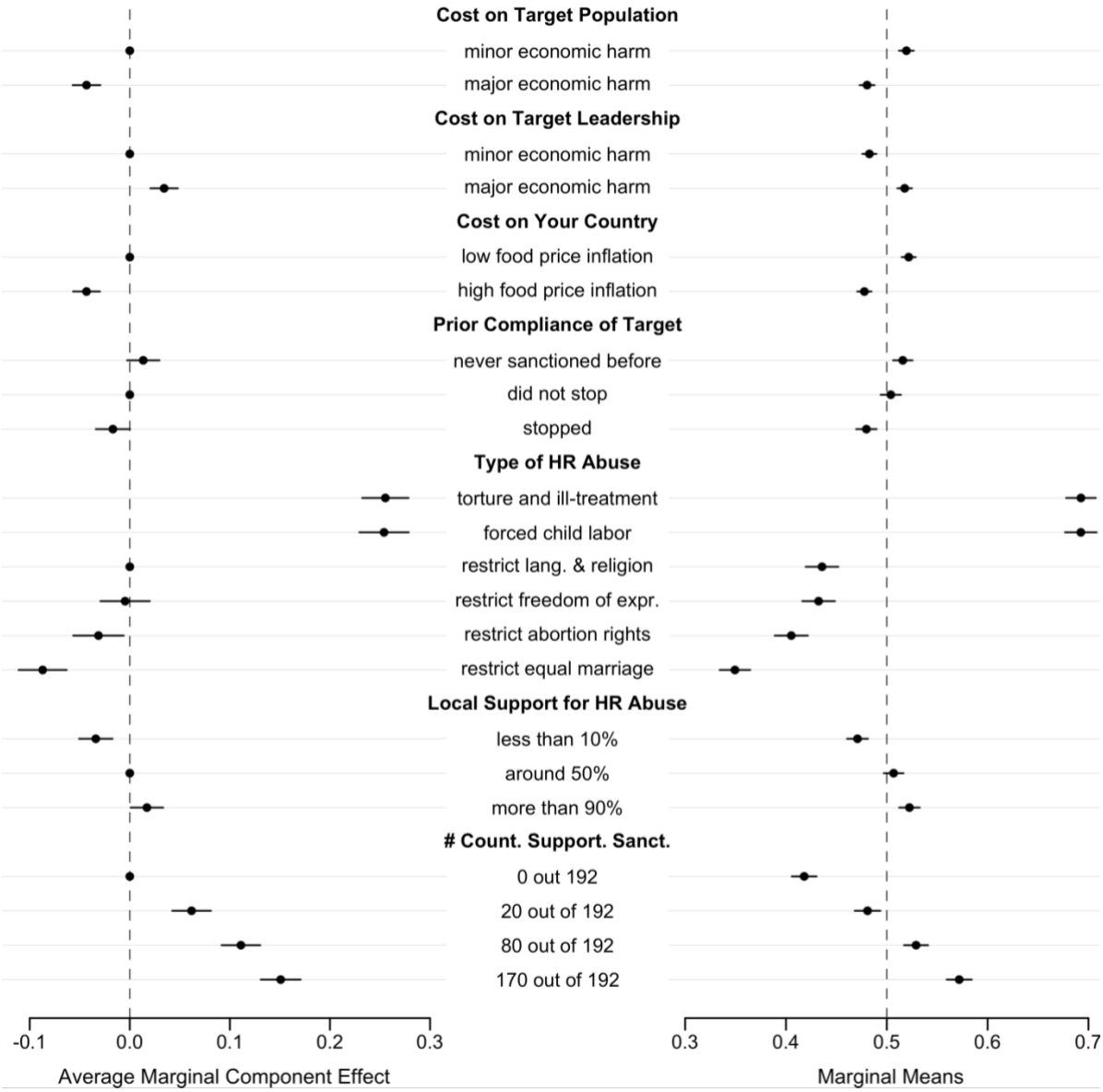


Figure 2. *Estimated Average Marginal Component Effects and Marginal Means. Dots represent point estimates and segments represent their 95% confidence intervals. Robust standard errors are clustered at the respondent level.* $N_{individuals} = 2001$ *and* $N_{observations} = 20010$.

Results also reveal that the type of human rights abuse does matter and that there is a hierarchy among different types of offenses in terms of their perceived salience to merit sanctions. *Forced child labor* and *Torture and ill-treatment by state authorities* are two categories that muster the highest support for sanctions. Compared to the baseline category of *Restricting language and religious practices of minorities*, respondents prefer imposing sanctions to these two high-offending categories by 25 percentage points ($p < 0.001$). *Restricting the freedom of expression* has no difference to the baseline category

whereas *Restricting abortion rights* and *Restricting equal marriage rights* are offenses that warrant the sanctions the least, respectively. In short, six distinct types of offenses form three clusters based on their perceived potency to merit sanctions. These findings are in line with Hypotheses 1.1 and 1.2.

We also investigate how respondent preferences vary with ideology and sex (see Appendix A2.2). Liberal/left-wing respondents are more willing to protect women's reproductive rights than conservative/right-wing individuals. That said, restricting abortion rights still does not have strong merit for imposing sanctions among liberal/left-wing individuals. The results also demonstrate that there is no meaningful difference in the effects of other types of violations by the respondent's political ideology. Thus, we do not find convincing evidence for Hypothesis 1.3. Turning to Hypothesis 1.4, our results indicate that neither men nor women prioritize issuing sanctions for abortion rights and that the level of support is not statistically different among the two groups. Indeed, the only meaningful difference we find between the two genders is that male respondents are more willing to impose sanctions against violations in relation to the freedom of expression. In short, we do not find evidence for Hypothesis 1.4.

Next, we turn to conditional cooperation and find that respondent support for sanctions monotonically increase as the number of countries involved in imposing sanctions rise. More precisely, relative to no support from other states at all, the probability of imposing sanctions is 6 percentage points higher when 20 foreign countries are also involved in the action and this difference raises to 15 percentage points if the sanction is supported by the majority of countries ($p < 0.001$). Taken together, our results demonstrate clear evidence in support of Hypothesis 6.1 that conditional cooperation is a salient factor in mustering public support sanctions against human rights violators.

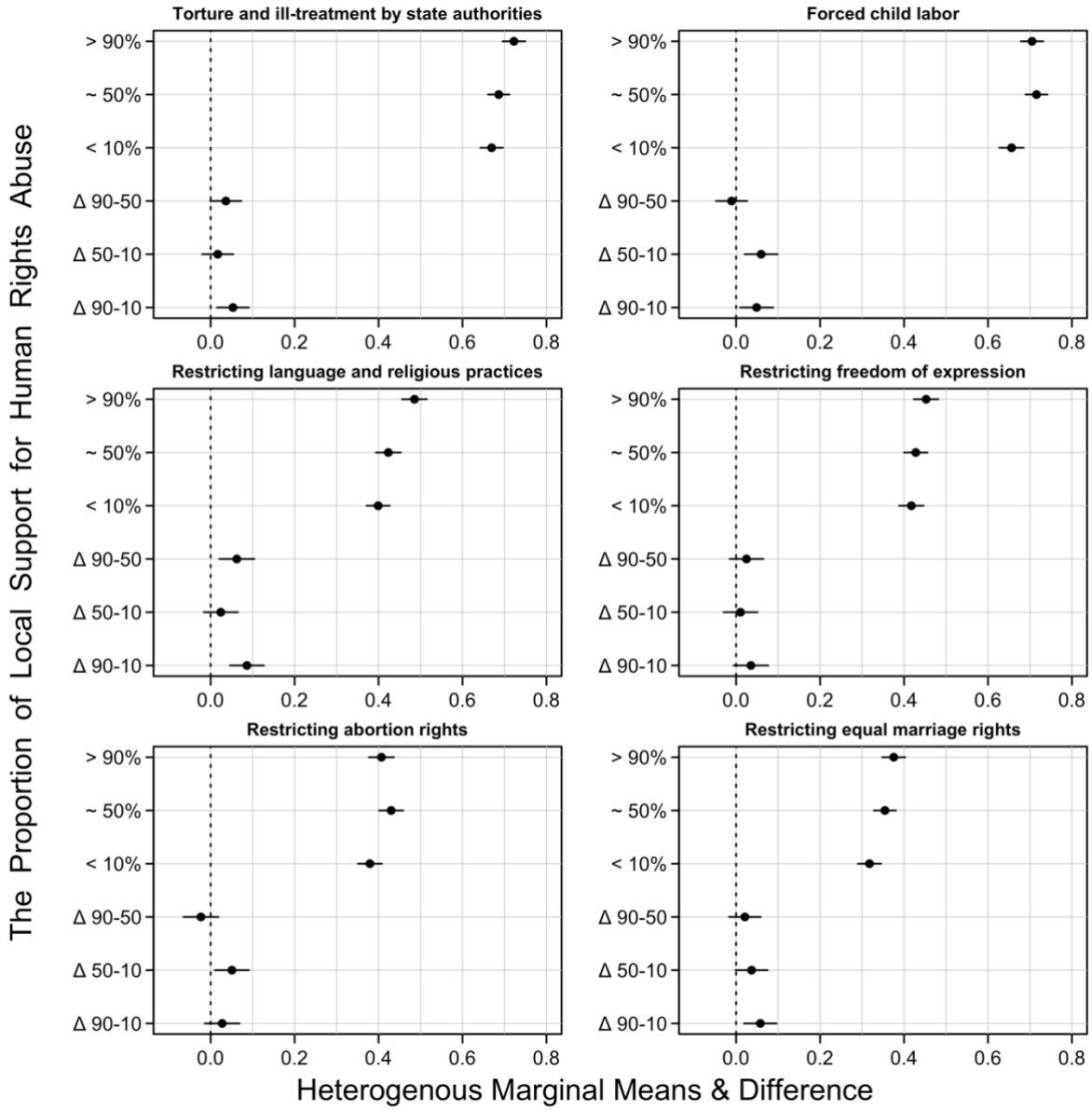


Figure 3. *Average Effects of Type of Human rights Abuse by Proportion of Local Support*

Last but not least, Figure 2 also illustrates that the proportion of local public support for the human rights abuse significantly affects the probability of supporting sanctions. Relative to the polarized condition (i.e., *around 50%*), respondents are less likely to impose sanctions if the proportion of local support for the human rights abuse is less than 10% but more likely to do so if the local support is more than 90%, by margins of 2 and 3 percentage points, respectively. However, the within-design interactions show that this relationship is conditional on a number of factors.

First, respondents are more willing to impose sanctions against two specific types of abuses, namely *Torture and ill-treatment by state authorities* (top left panel of Figure 3) and *Restricting language and religious practices of minorities* (middle left panel of Figure 3) once the majority of locals support the human rights abusing policies of their government. This result partially confirms Hypothesis 5.1 because increasing the local support toward *Restricting freedom of expression* does not affect respondents' willingness to impose sanctions. More interestingly, our results do not substantially support Hypothesis 5.2 that people disfavor sanctioning against contentious human rights abuses such as *Restricting abortion rights* and *Restricting equal marriage rights* even when the majority of the target population support the violating policy (bottom panels of Figure 3). Rather, we find that increasing the local support for both types of violations can also lead respondents to further skew their support for imposing sanctions. Respondents are more likely to play the norm entrepreneur role by which they are willing to forcibly alter the moral domain of the target.

Next, we turn to the interaction between local support for human rights infringing policy and the costs of sanctions on the target population (Figure 4). In line with

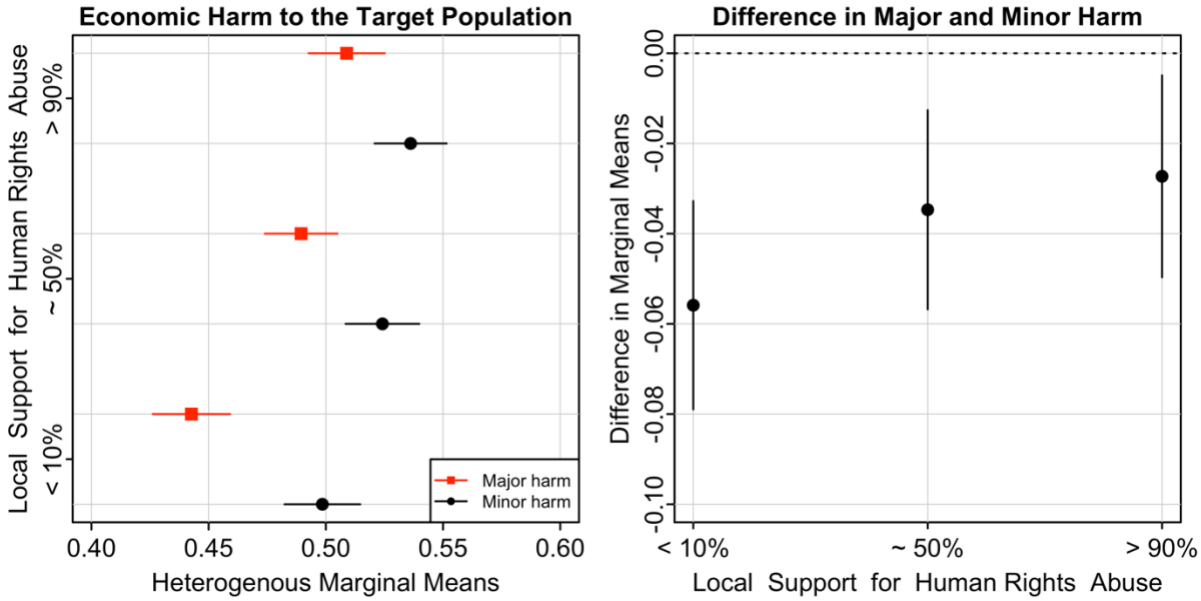


Figure 4. *Average Effect of Cost on the Target Population by Proportion of Local Support*

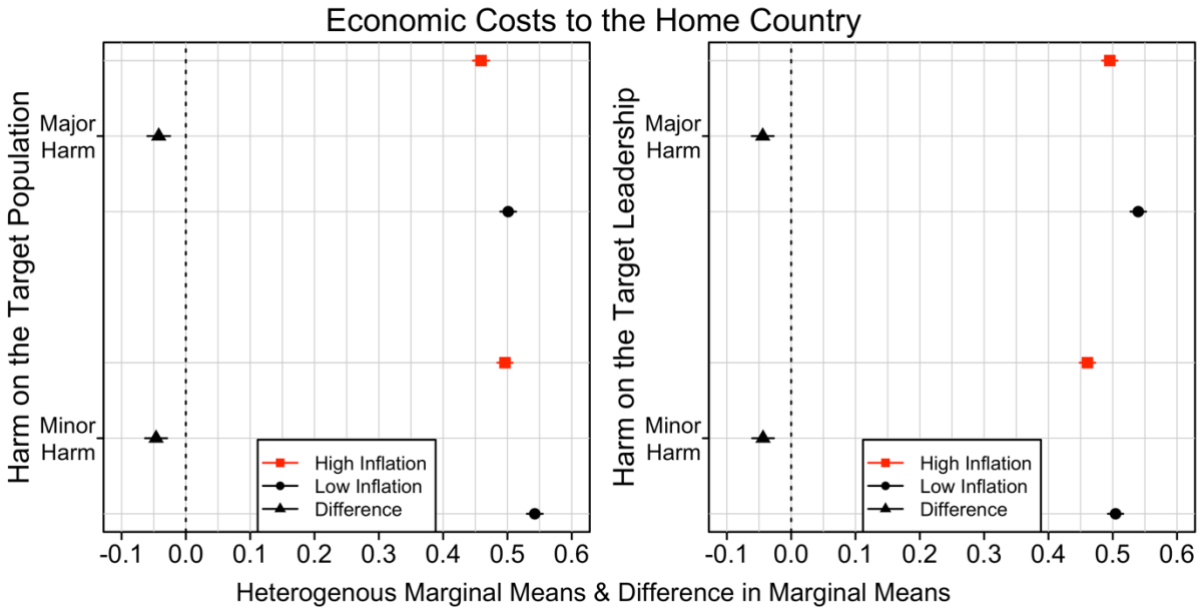


Figure 5. *Average Effect of Incurred Cost by Exported Cost*

Hypothesis 5.3, we find that the degree to which respondents disapprove of inflicting costs on the target population is conditional on the local support for human rights abuse. As the proportion of locals supporting the abusing policy increase, respondent willingness to shelter the target population from the costs of sanctions decrease. When the overwhelming majority of locals disapprove of their government's human rights abuse (i.e., *< 10%*), respondents further differentiate the target population and disfavor inflicting major harm by 6 percentage points. On the other hand, when the overwhelming majority of locals support the offending policy (i.e., *> 90%*), respondents disfavor major economic harm only by 3 percentage points. The difference between these two figures is significant at the 90% confidence level (see Appendix A2.3 for simulated coefficients). In short, respondents become less inclined to take the costs of sanctions on the target population into account when the level of support for the offending policies is high among the target population.

In contrast, we do not find a heterogenous treatment effect of incurred costs with respect to the harm falling on the target population. Contrary to Hypothesis 3.3,

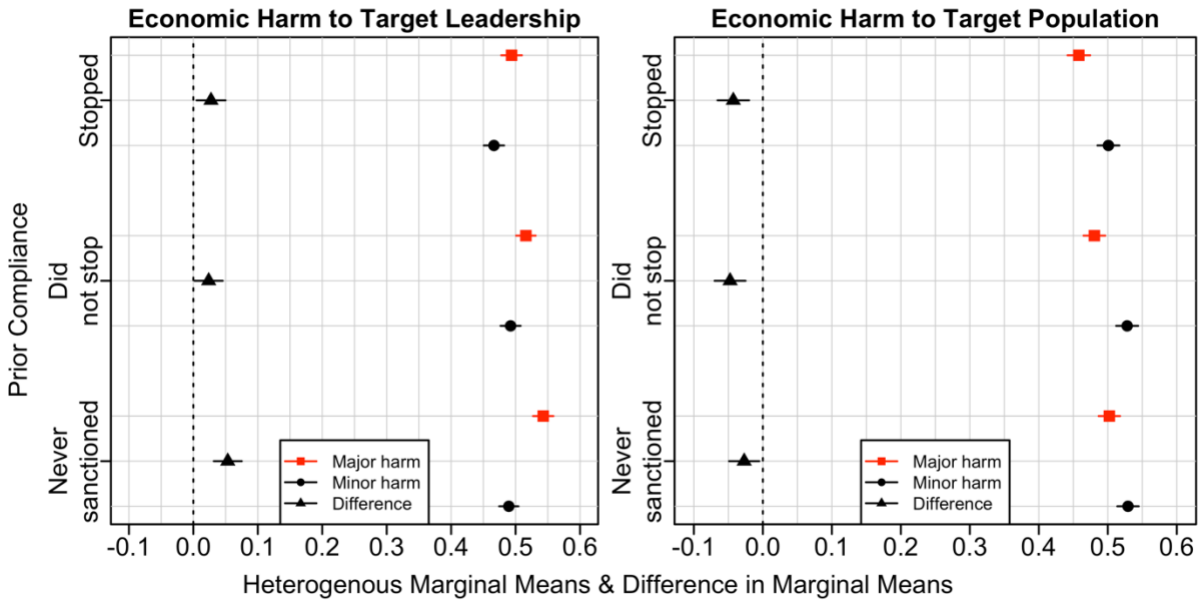


Figure 6. *Average Effect of Exported Cost on Target Population by Prior Compliance*

individuals do not become more willing to incur costs as the harm of sanctions on the target population decrease (left panel of Figure 5). Instead, the salience of incurred costs remains roughly the same regardless of the harm to the target population. When faced with exporting major and minor costs to the target population, respectively, respondents are 4.22 and 4.63 percentage points less likely to support sanctions that would cause higher inflation at home and this difference is insignificant. For completeness, we also investigate the exported cost on the target leadership (right panel of Figure 5) and find that the treatment effect is consistent across all groups. The probability of supporting sanctions that would cause higher inflation at home are 4.43 and 4.37 percentage points lower for major and minor costs of sanctions on the leadership, respectively, and this difference is again insignificant.

Although we find no evidence in support of Hypothesis 3.3, this null result is nonetheless indicative of respondent preferences, especially when juxtaposed against the findings regarding the interaction between local support and the cost of sanction on the target population. When faced with incurring higher costs, individuals remain consistent with their aversion irrespective of the context. Yet, they do consider contextual factors

when exporting costs. Based on the proportion of local support for government transgression, respondents adjust their disinclination toward inflicting harm on the target population.

Finally, we do not find support for Hypothesis 4.1 and 4.2, which formulate an interaction between prior compliance with sanctions and differential harm of exported costs (Figure 6). Respondents disregard the effectiveness of prior sanctions when considering inflicting harm to the target population or leadership. They remain consistent with their penchant to punish the leadership and with their aversion to harming the population, irrespective of the prior compliance. Again, the expressive function of sanctions in reinforcing morality and the desire to do something against the offenders may shed light on this disregard toward the effectiveness of prior sanctions.

Yet, the results suggest that respondents do consider the novelty associated with maiden sanctions. Compared to repeat sanctions, individuals are more likely to support maiden sanctions that would inflict major harm to the target leadership (left panel of Figure 6). The difference between major and minor harm on the leadership is 5.35 percentage points for the *never sanctioned* category whereas the same difference is 2.38 and 2.74 percentage points for the *did not stop* and *stopped* categories, respectively. This separation of the maiden sanctions from the repeat sanctions is statistically significant at the 90% confidence level. Similarly, the aversion toward imposing higher costs to the target population declines for the *never sanctioned* category (right panel of Figure 6). The difference between major and minor harm on the target population is -2.70 percentage points for the *never sanctioned* category whereas the same difference is -4.75 and -4.28 percentage points for the *did not stop* and *stopped* categories, respectively, but this separation is not statistically significant.

Contrary to our initial expectations but consistent with the rest of our findings, these results overlap with the "do something" argument, which emphasizes the expressive function of sanctions irrespective of their ability to influence policy outcomes. The novelty associated with maiden sanctions and the desire to do something against the offenders are likely reasons behind this heterogeneous treatment effect.

Robustness checks

We evaluate the robustness of our main results in multiple ways. We start with reporting the effect-consistency within countries by estimating our main model for each country separately (see Appendix A3.1). The findings suggest that the main effects of conjoint features remain consistent except the effect of type of human rights abuse. That is, respondents from the UK are more likely to impose sanctions against forced child labor and torture and ill-treatment by 31 percentage points ($p < 0.001$) relative to restricting language and religious practices, compared to the respondents from the U.S. by 20 percentage points ($p < 0.001$). In addition, American respondents are less likely to sanction against the violations of equal marriage rights and abortion rights than their UK counterparts by 13 and 9 percentage points ($p < 0.001$), respectively. However, the overall directions of findings in relation to our hypotheses overlap with the main results with the pooled sample.

Next, we evaluate the sensitivity of our findings to the forced-choice design by using an alternative seven-point scale variable as the outcome: "If you could vote on each of the countries in a referendum, how likely is it that you would vote in favor of or against imposing sanctions?", where higher values indicate stronger support for sanctions. The results verify that the conjoint features have similar effects on imposing sanctions (see Appendix A3.2).

We also investigate whether respondents' prior human rights attitudes and commitment influence our results. To this end, we use the 5-item human rights attitudes scale (Cohrs et al., 2007) to measure respondents' human rights commitment and categorize them into low and high commitment groups (*Cronbach's alpha > 0.70*). Following Stellmacher et al. (2005), we also use three other indicators of human rights commitment: donating money to a humanitarian organization; signing petitions against human rights violations; protests against human rights violations. Our main findings remain robust to such differentiation of sub-groups (see Appendix A3.3). For example, individuals are neither more tolerant toward the costs of sanctions nor more disapproving of economic harm on the target population than their low-commitment counterparts. In line with our main results, these again reflect the importance of sanction costs and consistency in respondent preferences toward a higher price tag.

It is also noteworthy that some respondents may be skeptical of their influence on foreign policy as mere citizens, thus we test whether the main findings are robust to such respondent perceptions and confirm that the results remain substantively comparable without significant heterogeneity (see Appendix A3.3). The only meaningful difference is that those people who have less confidence in their influence in policymaking are even more unwilling to issue sanctions against the violations of equal marriage rights. Moreover, varying perceptions regarding the recognition of human rights may obscure heterogeneity in the main effects, but further analysis confirms that those who believe that the majority of their fellow citizens support the protection of human rights have preferences similar to those who do not share such a conviction.

Finally, we test whether our experimental design holds its assumptions to ensure that inferences about the causal effects are credible (see Appendix A3.4). We first ensure that there are no *carryover effects* for sanctioning decisions. In other words, randomized

features of the country profiles in previous rounds do not significantly affect respondent decisions in the current round. Second, we confirm that the estimated effects of country profiles do not systemically vary depending on the profile's positionality in the conjoint tables. Lastly, we evaluate the extent to which each level of conjoint attributes is uniformly distributed and verify that any imbalance does not influence our estimates.

Conclusion

Publicly stressing the legal concepts of indivisibility and universality of human rights, democracies often reaffirm their continued commitment to protecting human rights abroad regardless of whoever the perpetrators are and wherever the abuses occur.[4] Yet, international responses to human rights violations do differ depending on the perpetrators and the type and perceived severity of the abuse. Human rights are indivisible in principle, but in practice, some violations are perceived as more atrocious than others. Our study presents evidence that such practical considerations and perceived hierarchies sway public opinion. This finding provides insights for uncovering why the principle of indivisibility of human rights does not immediately translate into sanction policies. The impulse to do something against the perpetrators varies in intensity according to the type and perceived severity of the abuse. In this respect, avowed 'redlines' in international human rights protection is not a mere signal of resolve to potential abusers, but also function as a reassertion reflecting the moral convictions of the public. Thus, our findings are also relevant for campaigners aiming to instigate action against human rights violators because how the violation is framed plays a key role in mustering public support.

[4] One example is the EU, which recently reasserted its conviction that human rights are indivisible (European Union 2021).

Sanctions regimes of the U.S. and UK, among other democracies, increasingly move away from conventional approaches of state-wide punishment and evolve toward targeting human rights abusers directly as a desirable and effective way of protecting the civilian population. While introducing UK's first autonomous human rights sanctions regime, Foreign Secretary Dominic Raab emphasized that the new legislation would allow the government "to target perpetrators without punishing the wider people of a country that may be affected" and further stressed the cross-party support for such targeted sanctions (2020). Our findings indicate an overlap between the emphasis the sanction senders put on protecting the target population and public sentiment at home. Yet, there exists a tension between the impulse to punish the perpetrators and the concern to shelter the target population. However, well-designed sanctions have unintended consequences of hurting the target population. Our results reveal that the public becomes more condoning toward these unintended consequences if the perpetrators enjoy support from their local population or the victims belong to minority groups. To gather the support of the public, sanction policies should be extra cautious not to harm the target population when the perpetrators are clearly acting in isolation in their transgressions.

Although respondents adjust their disapproval of inflicting harm on the target population based on the proportion of local support for the government transgression, their aversion toward the costs of sanctions remain remarkably consistent, indicating a lack of willingness to bear further costs for sheltering the target population. The absence of such a trade-off between the moral and the instrumental reveals the salience of self-centered considerations. On the other hand, sanctions do play a key role in satisfying the impulse to do something against the perpetrators, as shown by the stronger support for maiden sanctions. Thus, emphasizing the novelty of proposed sanctions is an effective strategy to inspire the backing of the public.

Similarly, multilateral coordination is a crucial factor in communicating to the public that sanctions are indeed necessary and the right policy to deploy. Democracies often declare sanctions on human rights violators in unison. For instance, U.S., UK, and Canada, announced sanctions against the Myanmar military on 10 December 2021 as a response to human rights violations and abuses (Foreign Office 2021). Our results indicate that announcing sanctions alongside other countries is effective at conveying the message that the sanctions are deployed as a vital response. The experimental evidence is clear that the public support for deploying sanctions gets stronger as the number of third-party countries involved in sanctions increases. Thus, failing to participate in multilateral sanctions is likely to draw the disapproval of the public, a point that human rights organizations and campaigners can utilize to pressure governments not to condone the perpetrators.

Finally, we identify avenues for further research. Studies investigating the micro-foundations of public attitudes so far isolated sanctions as a stand-alone policy, but further research should progress toward analyzing the integration of sanctions to general diplomacy and unpack the connections between sanctions and other foreign policy instruments. Considering that sanctions against human rights abusers are often deployed as a response short of military action, studying humanitarian intervention vis-à-vis economic sanctions is likely to shed further light on the predicaments emerging from navigating the tradeoffs between instrumental and moral considerations. In this respect, the harm falling on the target population is a crucial dimension worth further investigation because even the well-intended and the best crafted humanitarian intervention would inevitably cause some unintended consequences of harming the civilian population.

References

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2021. 'Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments'. *Political Science Research and Methods* 9 (1): 53–71.

BBC. 2014. 'US Imposes Sanctions on Uganda for Anti-Gay Law'. BBC News. 19 June 2014. https://www.bbc.co.uk/news/world-us-canada-27933051.

Bechtel, Michael M., and Kenneth F. Scheve. 2013. 'Mass Support for Global Climate Agreements Depends on Institutional Design'. *Proceedings of the National Academy of Sciences of the United States of America* 110 (34): 13763–68.

Cheung, Stephen L. 2014. 'New Insights into Conditional Cooperation and Punishment from a Strategy Method Experiment'. *Experimental Economics* 17 (1): 129–53.

Chilton, Adam 2014. 'The Influence of International Human Rights Agreements on Public Opinion'. *Chicago Journal of International Law* 15 (1): 110–37.

———. 2015. 'The Laws of War and Public Opinion: An Experimental Study'. *Journal of Institutional and Theoretical Economics* 171 (1): 181–201.

Chilton, Adam, and Mila Versteeg. 2016. 'Do Constitutional Rights Make a Difference?' *American Journal of Political Science* 60 (3): 575–89.

Christiansen, William, Tobias Heinrich, and Timothy M. Peterson. 2016. 'Backyard Effects in Democracy Promotion: Evidence from a Survey Experiment'.

Cohrs, J. Christopher, Jürgen Maes, Barbara Moschner, and Sven Kielmann. 2007. 'Determinants of Human Rights Attitudes and Behavior: A Comparison and Integration of Psychological Perspectives'. *Political Psychology* 28 (4): 441–69.

Conrad, Courtenay R., Daniel W. Hill, and Will H. Moore. 2018. 'Torture and the Limits of Democratic Institutions'. *Journal of Peace Research* 55 (1): 3–17.

Drezner, Daniel W. 2011. 'Sanctions Sometimes Smart: Targeted Sanctions in Theory and Practice'. *International Studies Review* 13 (1): 96–108.

———. 2015. 'Targeted Sanctions in a World of Global Finance'. *International Interactions* 41 (4): 755–64.

Druckman, James N., ed. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge ; New York: Cambridge University Press.

European Union. 2021. 'Declaration by the High Representative Josep Borrell'. European Council. 9 December 2021. https://www.consilium.europa.eu/en/press/press-releases/2021/12/09/human-rights-day-declaration-by-the-high-representative-josep-borrell-on-behalf-of-the-european-union/.

Finnemore, Martha. 1996. 'Constructing Norms of Humanitarian Intervention'. In *The Culture of International Security*, edited by Peter Katzenstein, 153–85. New York: Columbia University Press.

Finnemore, Martha, and Kathryn Sikkink. 1998. 'International Norm Dynamics and Political Change'. *International Organization*, 887–917.

Foreign Office. 2021. 'New UK Sanctions Target Human Rights Violations and Abuses in Myanmar and Pakistan'. UK Government Press Release. 10 December 2021. https://www.gov.uk/government/news/new-uk-sanctions-target-human-rights-violations-and-abuses-in-myanmar-and-pakistan.

Galtung, Johan. 1967. 'On the Effects of International Economic Sanctions, With Examples from the Case of Rhodesia'. *World Politics* 19 (3): 378–416.

Gartner, Scott. 2008. 'The Multiple Effects of Casualties on Public Support for War: An Experimental Approach'. *American Political Science Review* 102 (1): 95–106.

Goldenberg, Suzanne. 2007. 'Bush Orders Darfur Sanctions after Grassroots Pressure'. *The Guardian*, 30 May 2007. https://www.theguardian.com/world/2007/may/30/usa.sudan.

Green, Paul., Abba Krieger, and Yoram Wind. 2001. 'Thirty Years of Conjoint Analysis: Reflections and Prospects'. *INFORMS Journal on Applied Analytics* 31 (3): S56–57.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. 'Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments'. *Political Analysis* 22 (1): 1–30.

Hansen, Kasper M., Asmus L. Olsen, and Mickael Bech. 2015. 'Cross-National Yardstick Comparisons: A Choice Experiment on a Forgotten Voter Heuristic'. *Political Behavior* 37 (4): 767–89.

Heinrich, Tobias, and Yoshiharu Kobayashi. 2020. 'How Do People Evaluate Foreign Aid To "Nasty" Regimes?' *British Journal of Political Science* 50 (1): 103–27.

Heinrich, Tobias, Yoshiharu Kobayashi, and Leah Long. 2018. 'Voters Get What They Want (When They Pay Attention): Human Rights, Policy Benefits, and Foreign Aid'. *International Studies Quarterly* 62 (1): 195–207.

Heinrich, Tobias, Yoshiharu Kobayashi, and Timothy M. Peterson. 2017. 'Sanction Consequences and Citizen Support: A Survey Experiment'. *International Studies Quarterly* 61 (1): 98–106.

Human Rights Watch. 2019. 'Human Rights Watch Country Profiles: Sexual Orientation and Gender Identity'. Human Rights Watch World Report 2019. 23 September

2019. https://www.hrw.org/video-photos/interactive/2019/02/28/human-rights-watch-country-profiles-sexual-orientation-and.

Hyde, Susan D. 2011. 'Catch Us If You Can: Election Monitoring and International Norm Diffusion'. *American Journal of Political Science* 55 (2): 356–69.

Kirshner, Jonathan. 1997. 'The Microfoundations of Economic Sanctions'. *Security Studies* 6 (3): 32–64.

Kreps, Sarah, and Sarah Maxey. 2018. 'Mechanisms of Morality: Sources of Support for Humanitarian Intervention'. *Journal of Conflict Resolution* 62 (8): 1814–42.

Kustra, Tyler. 2021. 'Sanctioning the Homeland: Diasporas' Influence on American Economic Sanctions Policy'. *Journal of Conflict Resolution*, November, 002200272110426.

Leeper, Thomas J., Sara B. Hobolt, and James Tilley. 2020. 'Measuring Subgroup Preferences in Conjoint Experiments'. *Political Analysis* 28 (2): 207–21.

Malloy, Michael P. 1995. 'Economic Sanctions and Human Rights: A Delicate Balance' 3 (1): 1–2.

McFarland, Sam, and Melissa Mathews. 2005. 'Who Cares about Human Rights?' *Political Psychology* 26 (3): 365–85.

McLean, Elena V., and Taehee Whang. 2014. 'Designing Foreign Policy: Voters, Special Interest Groups, and Economic Sanctions'. *Journal of Peace Research* 51 (5): 589–602.

Morgan, T. Clifton, and Valerie L. Schwebach. 1997. 'Fools Suffer Gladly: The Use of Economic Sanctions in International Crises'. *International Studies Quarterly* 41 (1): 27–50.

Murdie, Amanda, and Dursun Peksen. 2013. 'The Impact of Human Rights INGO Activities on Economic Sanctions'. *Review of International Organizations* 8 (1): 33–53.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. 'Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research'. *Journal of Experimental Social Psychology* 70 (May): 153–63.

Peksen, Dursun. 2009. 'Better or Worse? The Effect of Economic Sanctions on Human Rights *'. *Journal of Peace Research* 46 (1): 59–77.

Peksen, Dursun, Timothy M. Peterson, and A. Cooper Drury. 2014. 'Media-Driven Humanitarianism? News Media Coverage of Human Rights Abuses and the Use of Economic Sanctions'. *International Studies Quarterly* 58 (4): 855–66.

Putnam, Tonya, and Jacob Shapiro. 2017. 'International Law and Voter Preferences: The Case of Foreign Human Rights Violations'. *Human Rights Review* 18 (3): 243–62.

Raab, Dominic. 2020. 'Global Human Rights Sanctions Regime: Foreign Secretary's Statement to Parliament'. Oral Statement to Parliament. 6 July 2020. https://www.gov.uk/government/speeches/statement-on-the-global-human-rights-sanctions-regime.

Stein, Rachel M. 2015. 'War and Revenge: Explaining Conflict Initiation by Democracies'. *American Political Science Review* 109 (3): 556–73.

Stellmacher, Jost, Gert Sommer, and Elmar Brahler. 2005. 'The Cognitive Representation of Human Rights: Knowledge, Importance, and Commitment'. *Peace and Conflict* 11 (3): 267–92.

Time. 1966. 'United Nations: Sanctions Against Rhodesia', 23 December 1966. http://content.time.com/time/subscriber/article/0,33009,840760,00.html.

Tomz, Michael R., and Jessica L. P. Weeks. 2020. 'Human Rights and Public Support for War'. *Journal of Politics* 82 (1): 182–94.

U.S. Congress. 1998. Subcommittee On International Economic Policy And Trade Of The House International Relations Committee Holds Hearing On The Successes And Failures Of U.s. Economic Engagement With China.

Weiss, Thomas G. 1999. 'Sanctions as a Foreign Policy Tool: Weighing Humanitarian Impulses'. *Journal of Peace Research* 36 (5): 499–509.

Whang, Taehee. 2011. 'Playing to the Home Crowd? Symbolic Use of Economic Sanctions in the United States'. *International Studies Quarterly* 55 (3): 787–801.

Wheeler, Nicholas. 2000. *Saving Strangers: Humanitarian Intervention in International Society*. Oxford University Press.

Whitmeyer, Joseph M. 2002. 'The Compliance You Need for a Cost You and Collective Sanctions?' *Social Science Research* 31: 630–52.

Wike, Richard, and Shannon Schumacher. 2020. 'Attitudes toward Democratic Rights and Institutions'. Pew Research Center. 27 February 2020. https://www.pewresearch.org/global/2020/02/27/attitudes-toward-democratic-rights-and-institutions/.

Wood, Reed M. 2008. '"A Hand upon the Throat of the Nation": Economic Sanctions and State Repression, 1976-2001'. *International Studies Quarterly* 52 (3): 489–513.

Appendix for

Human Rights Violations and Public Support for Sanctions

Appendix A: Experimental Design, Sample & Results

1. Survey Experiment and Sample Characteristics

The data in this research is based on a Web-based conjoint experiment embedded in an online survey that we conducted in July 2020. Our pooled dataset consists of 2001 respondents from the United Kingdom (N=1009) and the United States (N=992). Participants were recruited through Prolific's representative quota sampling that provided us with fairly representative and high-quality datasets of the U.S. population and the UK population based on age, sex, and ethnicity. Figure 7 demonstrates the instructions and paired conjoint profiles of human-rights abusing countries during the study. Figure 8 and Figure 9 also show the demographic characteristics of both the UK and U.S. datasets, respectively.

Please carefully review the information detailed below, then please answer the questions. Note that you will evaluate **a different pair of Country A and Country B in 5 rounds.**

	Country A	Country B
Type of Human Rights Abuse	Restricting language and religious practices of minorities	Forced child labor
The Percentage of Citizens in the Targeted Country Supporting the Abusive Policy	around 50%	around 50%
The Number of Countries Supporting the Sanction	20 out of 192	20 out of 192
Whether the Country Previously Stopped Abuse After Earlier Sanctions	Did not stop	Stopped
Cost of the Sanction on the UK	Low food price inflation	High food price inflation
Cost of the Sanction on the Leadership of Target	Major economic harm	Major economic harm
Cost of the Sanction on the General Population of Target	Major economic harm	Minor economic harm

Which of these countries would you like your government to impose economic sanctions the most?

◯ Country A

◯ Country B

Figure 7. *Conjoint Profiles*

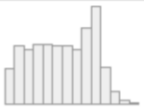
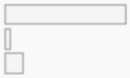
Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
sex [factor]	1. Male 2. Female	4940 (49.0%) 5150 (51.0%)		0 (0%)
age [numeric]	Mean (sd) : 45.6 (15.6) min < med < max: 18 < 46 < 82 IQR (CV) : 28 (0.3)	63 distinct values		0 (0%)
ethnic [factor]	1. White 2. Black 3. Other	8460 (83.9%) 390 (3.9%) 1240 (12.3%)		0 (0%)
politics [factor]	1. Liberal 2. Conservative 3. Other	4540 (45.0%) 2540 (25.2%) 3010 (29.8%)		0 (0%)
educ [factor]	1. No degree 2. Degree holder 3. Higher degree holder	4950 (49.1%) 3070 (30.4%) 2070 (20.5%)		0 (0%)
religion [factor]	1. No religion 2. Christian 3. Other	5380 (53.3%) 3860 (38.3%) 850 (8.4%)		0 (0%)

Figure 8. *UK Sample*

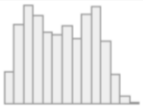
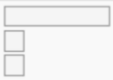
Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
sex [factor]	1. Male 2. Female	4840 (48.8%) 5080 (51.2%)		0 (0%)
age [numeric]	Mean (sd) : 45.4 (16) min < med < max: 18 < 45 < 84 IQR (CV) : 29 (0.4)	63 distinct values		0 (0%)
ethnic [factor]	1. White 2. Black 3. Other	7230 (72.9%) 1340 (13.5%) 1350 (13.6%)		0 (0%)
politics [factor]	1. Liberal 2. Conservative 3. Other	5940 (59.9%) 2400 (24.2%) 1580 (15.9%)		0 (0%)
educ [factor]	1. No degree 2. Degree holder 3. Higher degree holder	2610 (26.3%) 5140 (51.8%) 2170 (21.9%)		0 (0%)
religion [factor]	1. No religion 2. Christian 3. Other	4440 (44.8%) 4620 (46.6%) 860 (8.7%)		0 (0%)

Figure 9. *U.S. Sample*

1.2 Data Quality Checks

Figure 10 and Figure 11 illustrate the distributions of survey completion time across respondents who successfully comprehended the instructions and did not. We excluded those respondents who rushed through and failed in the attention check in relation to the study instructions from the final sample.

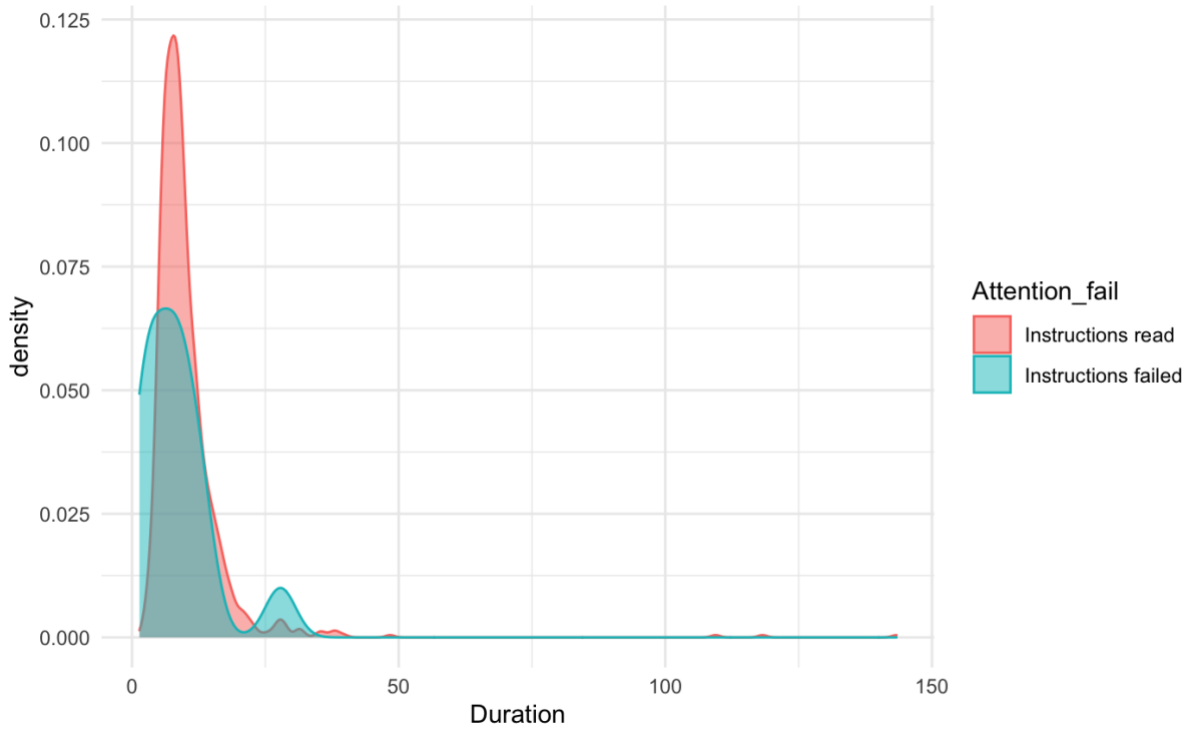


Figure 10. *Completion Time in Minutes (UK Sample)*

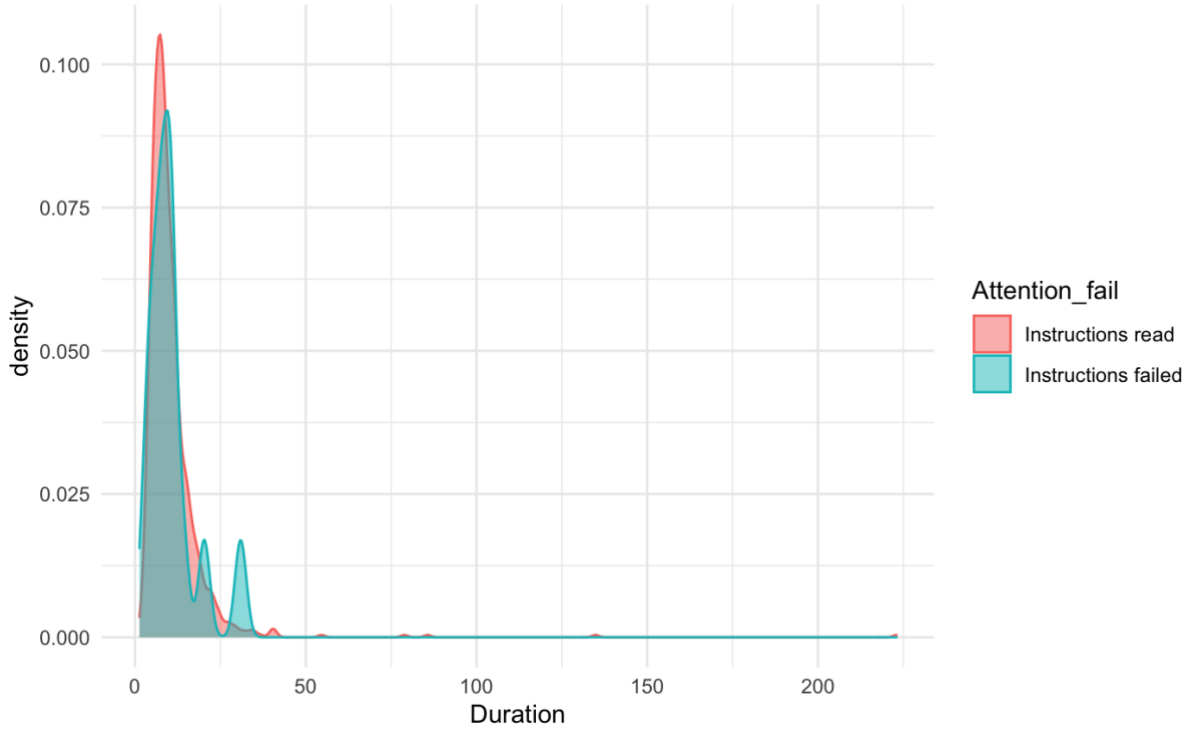


Figure 11. *Completion Time in Minutes (US Sample)*

1.3 Power Analysis

Statistical power in conjoint experiments is a function of the number of repeated trials performed by each respondent, the number of levels of an attribute, and the size of the measured effect in the population. Stefanelli and Lukac, (2020) show that effect sizes in conjoint experiments are often small despite large sample sizes. This tend to be the case because researchers either overload their design with a high number of experimental conditions or have limited trials. Reflecting on these concerns, we performed a *post hoc* power analysis to test whether our conjoint design is well-powered for our samples, using the power analysis tool proposed by Lukac and Stefanelli (2020). Given our design features, Figure 12 and Figure 13 show that both studies (UK sample and U.S. sample) satisfactorily reach the conventional power threshold (≥ 0.80).

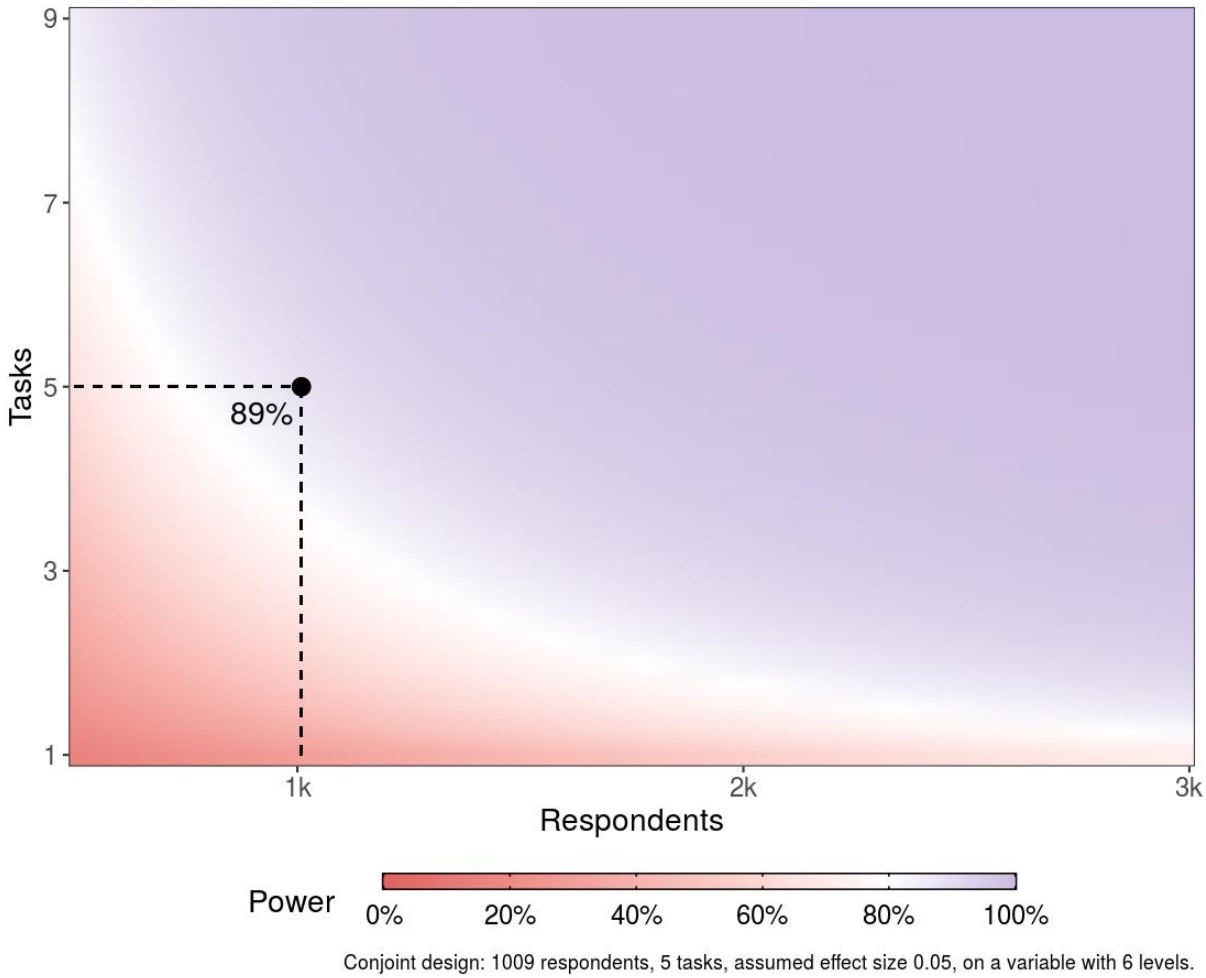


Figure 12. *Power Analysis (UK Sample)*

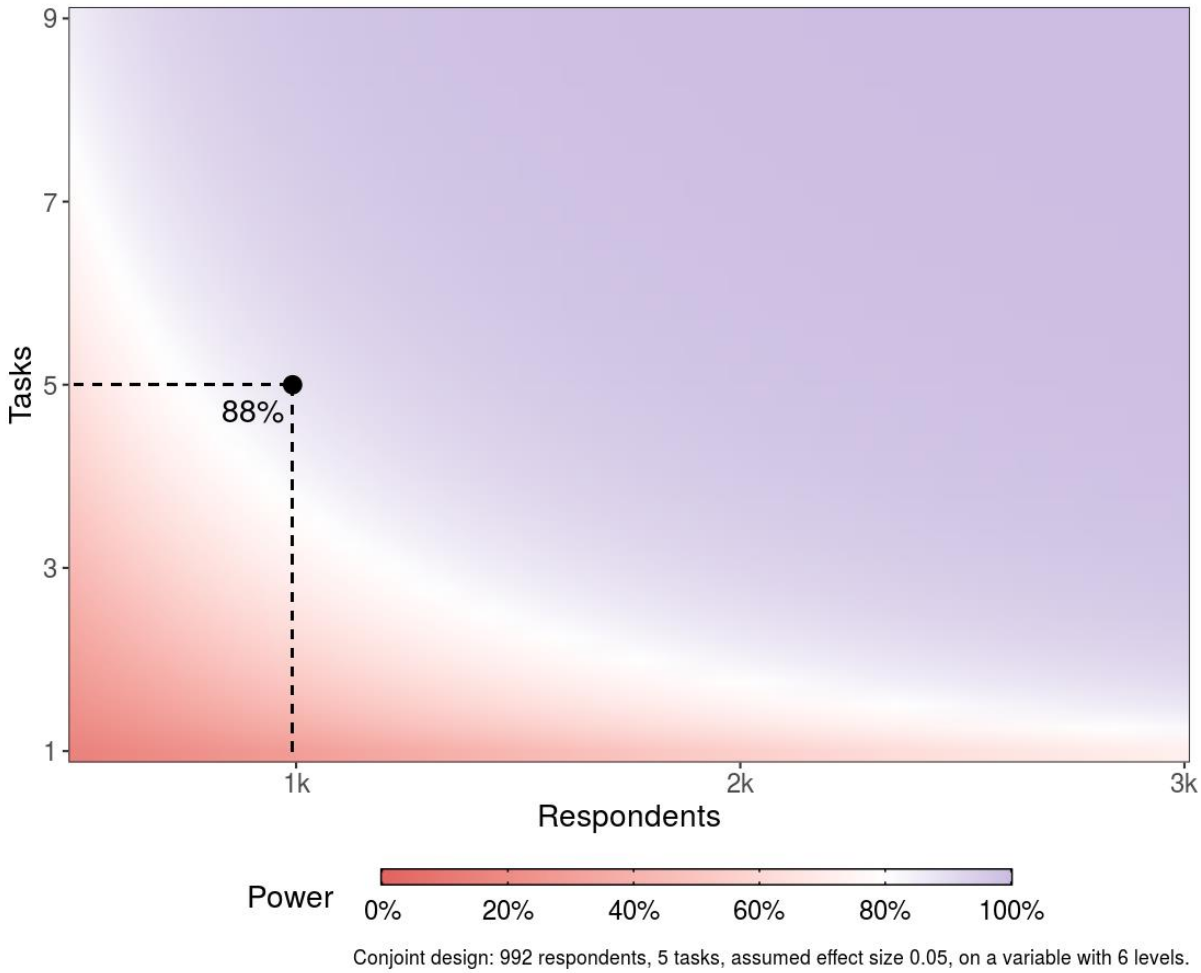


Figure 13. *Power Analysis (US Sample)*

2. Main Results

Table 2. Estimated Average Marginal Component Effects without Controls

Attribute	Level	Estimates	Std. Err.	P-Value
Cost_of_Sanction_on_General_Population_of_Target	major economic harm	-0.04	0.01	0.00
Cost_of_Sanction_on_Leadership_of_Target	Major economic harm	0.03	0.01	0.00
Cost_of_Sanction_on_Your_Country	High food price inflation	-0.04	0.01	0.00
Prior_Compliance_Behaviour_of_Target	Did not stop	-0.01	0.01	0.10
Prior_Compliance_Behaviour_of_Target	Stopped	-0.03	0.01	0.00
The_Number_of_Countries_Supporting_the_Sanction	20 out of 192	0.06	0.01	0.00
The_Number_of_Countries_Supporting_the_Sanction	80 out of 192	0.11	0.01	0.00
The_Number_of_Countries_Supporting_the_Sanction	170 out of 192	0.15	0.01	0.00
The_Proportion_of_Local_Human_Rights_Abuse_Support	more than 90%	0.02	0.01	0.04
The_Proportion_of_Local_Human_Rights_Abuse_Support	less than 10%	-0.03	0.01	0.00
Type_of_Human_Rights_Abuse	Restricting freedom of expression	0.00	0.01	0.71
Type_of_Human_Rights_Abuse	Forced child labor	0.25	0.01	0.00
Type_of_Human_Rights_Abuse	Restricting abortion rights	-0.03	0.01	0.02
Type_of_Human_Rights_Abuse	Restricting equal marriage rights	-0.09	0.01	0.00
Type_of_Human_Rights_Abuse	Torture and ill-treatment by state authorities	0.26	0.01	0.00

Table 3. Estimated Average Marginal Component Effects with Controls

Attribute	Level	Estimates	Std. Err.	P-Value
age_group	31-60	0.00	0.00	0.35
age_group	61+	0.01	0.00	0.03
Cost_of_Sanction_on_General_Population_of_Target	major economic harm	-0.04	0.01	0.00
Cost_of_Sanction_on_Leadership_of_Target	Major economic harm	0.03	0.01	0.00
Cost_of_Sanction_on_Your_Country	High food price inflation	-0.04	0.01	0.00
educ	Degree holder	0.00	0.00	0.70
educ	Higher degree holder	0.00	0.00	0.52
ethnic	Black	-0.01	0.00	0.01
ethnic	Other	0.00	0.00	0.32
Prior_Compliance_Behaviour_of_Target	Did not stop	-0.01	0.01	0.10
Prior_Compliance_Behaviour_of_Target	Stopped	-0.03	0.01	0.00
sex	Female	0.00	0.00	0.73
The_Number_of_Countries_Supporting_the_Sanction	20 out of 192	0.06	0.01	0.00
The_Number_of_Countries_Supporting_the_Sanction	80 out of 192	0.11	0.01	0.00
The_Number_of_Countries_Supporting_the_Sanction	170 out of 192	0.15	0.01	0.00
The_Proportion_of_Local_Human_Rights_Abuse_Support	more than 90%	0.02	0.01	0.04
The_Proportion_of_Local_Human_Rights_Abuse_Support	less than 10%	-0.03	0.01	0.00
Type_of_Human_Rights_Abuse	Restricting freedom of expression	0.00	0.01	0.71
Type_of_Human_Rights_Abuse	Forced child labor	0.25	0.01	0.00
Type_of_Human_Rights_Abuse	Restricting abortion rights	-0.03	0.01	0.02
Type_of_Human_Rights_Abuse	Restricting equal marriage rights	-0.09	0.01	0.00
Type_of_Human_Rights_Abuse	Torture and ill-treatment by state authorities	0.26	0.01	0.00

2.2 Ideology and Sex

Respondent preferences may vary with ideology (Figure 14) and sex (Figure 15). Although liberal/left-wing respondents are more willing to protect women's reproductive rights than conservative/right-wing individual, restricting abortion rights still does not have strong merit for imposing sanctions even among liberal/left-wing individuals. The results also demonstrate that there is no meaningful difference in the effects of other types of violations by the respondent's political ideology.

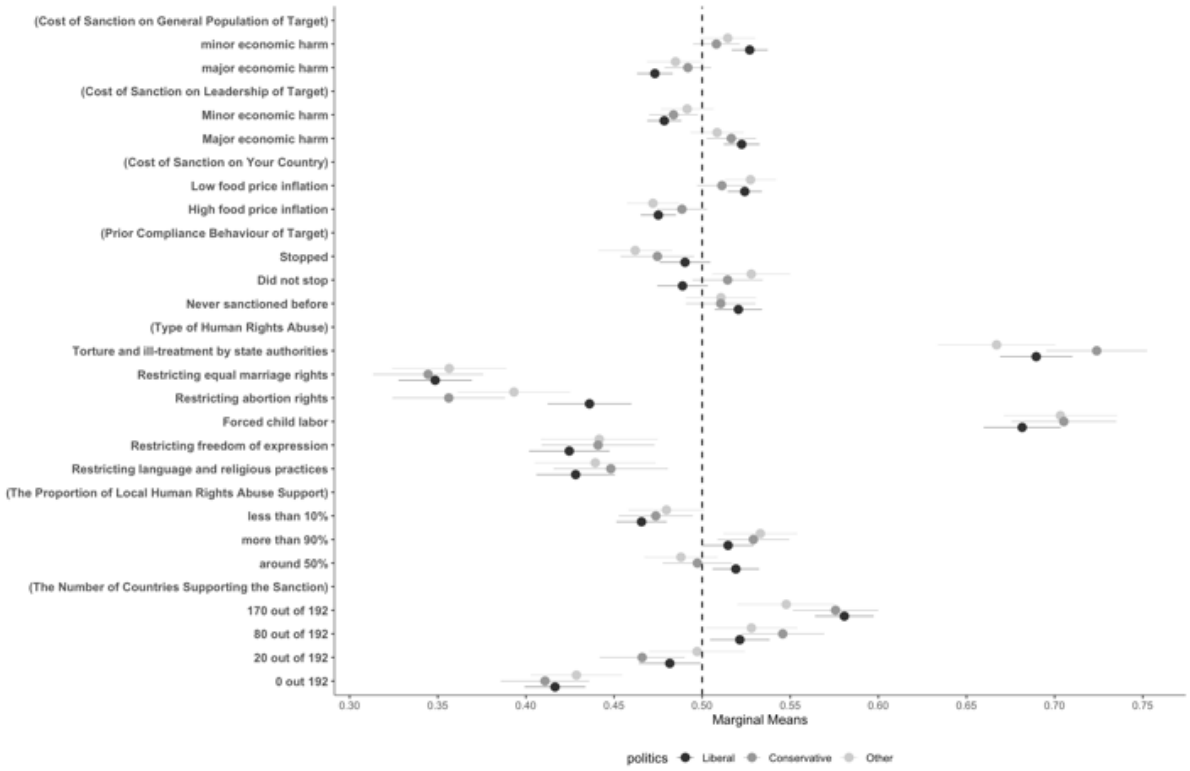


Figure 14. *Marginal Means by Political Ideology Notes: Dots represent point estimates of marginal means, and segments represent their 95% confidence intervals. Robust standard errors are clustered at respondent level.* $N_{individuals}$*=2001 &* $N_{observations}$*=20010. In our forced-choice conjoint design with two profiles per choice task, marginal means have a direct interpretation as probabilities: these MMs average 0.5 with values above 0.5 indicating features that increase sanction's favorability and values below 0.5 indicating features that decrease sanction's favorability.*

Turning to sex (Figure 15), neither men nor women prioritize issuing sanctions for abortion rights, and the level of support is not statistically different among the two groups. Indeed, the only meaningful difference is that female respondents are more discouraged to impose sanctions against violations in relation to the freedom of expression.

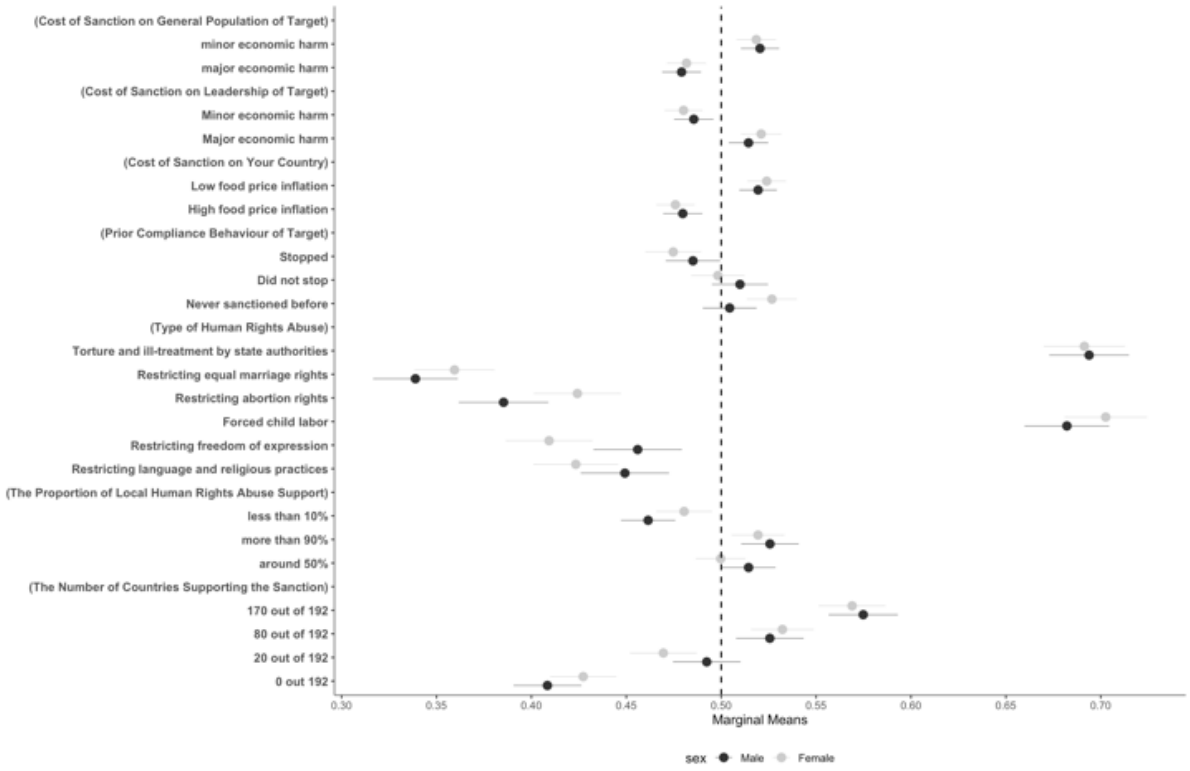


Figure 15. *Marginal Means by Sex.*

2.3 Simulated Coefficient Differences

Figures below show the differences between differences of marginal means. We first estimate the difference of marginal means between two categories for a specific treatment condition X (giving μ_X and σ_X). Next, we estimate the same difference for a second treatment condition Y (giving μ_Y and σ_Y). In total, we have two differences of marginal means. We then compare the changes in marginal means for two respective conditions by drawing coefficients from a multivariate normal distribution (N = 10,000) where;

$$\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \; ; \; \sigma = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$$

We then calculate the differences of change in respective marginal means $\mu_X - \mu_y$. Finally, we report differences in percentage points.

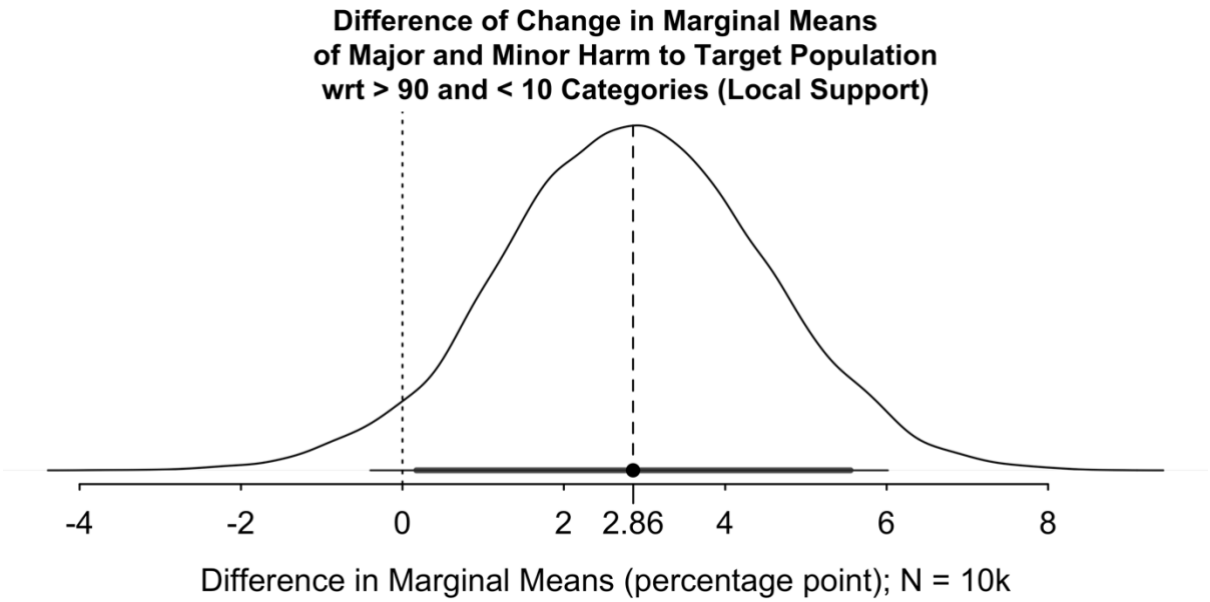


Figure 16. *This figure compares the percentage point changes in marginal means from major to minor harm (target population) with respect to high (90%) and low (10%) support for violations from the target population. Thick and thin lines represent 90% and 95% confidence intervals, respectively. See Figure 4 in main document.*

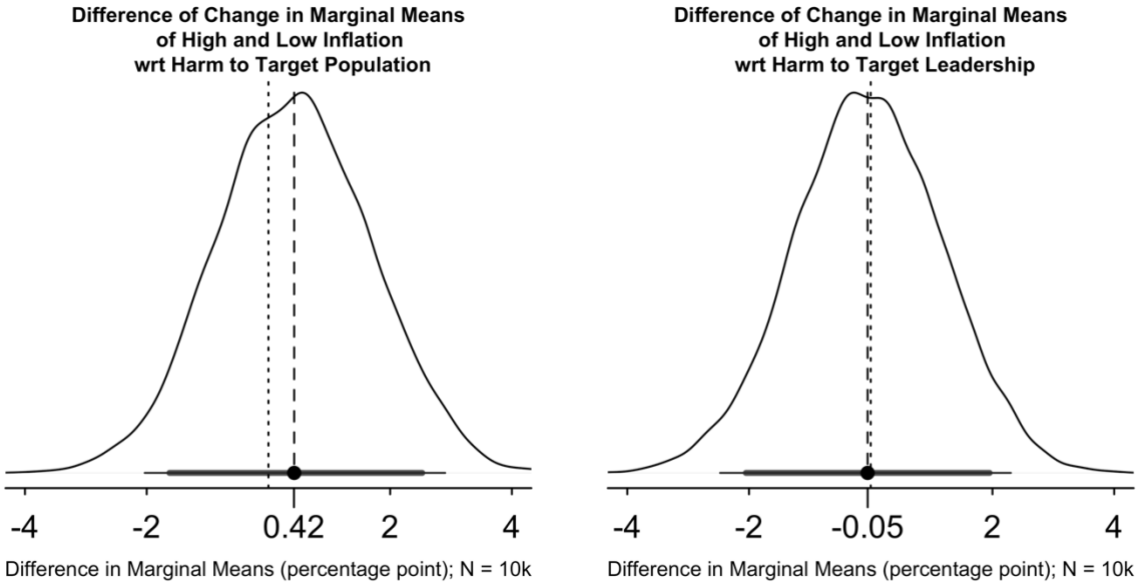


Figure 17. *This figure compares the changes in marginal means from high inflation to low inflation with respect to major and minor harm to target population (left panel) and leadership (right panel). Thick and thin lines represent 90% and 95% confidence intervals, respectively. See Figure 5 in main document.*

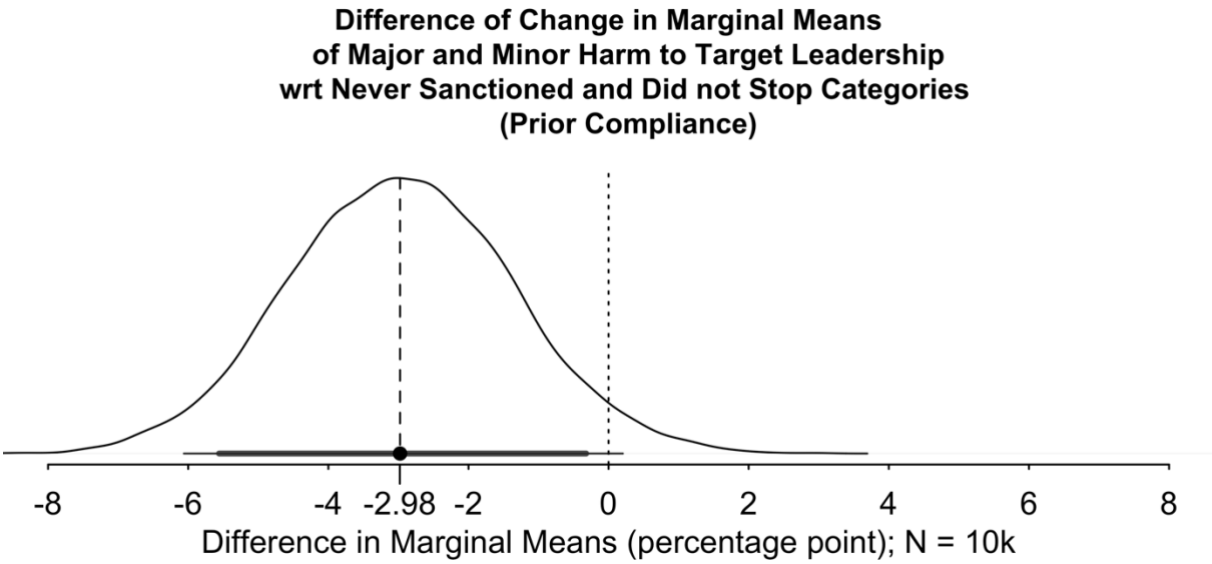


Figure 18. *This figure compares the percentage point changes in marginal means from major to minor harm (target leadership) with respect to never sanctioned and did not stop the abuse categories (prior compliance). Thick and thin lines represent 90% and 95% confidence intervals, respectively. See Figure 6 in main document.*

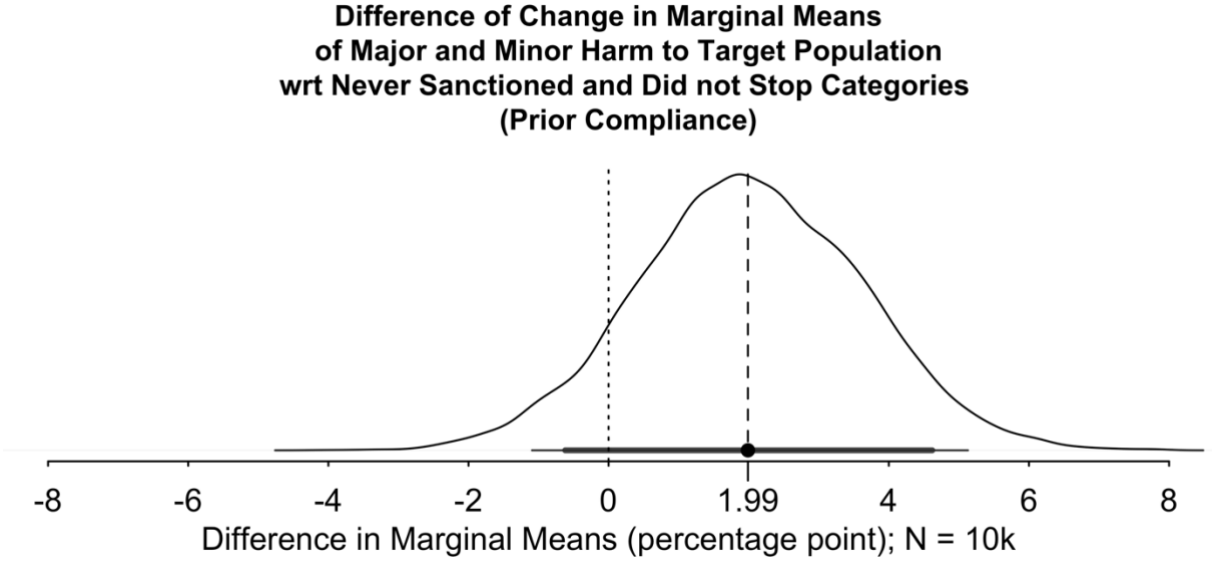


Figure 19. *This figure compares the percentage point changes in marginal means from major to minor harm (target population) with respect to never sanctioned and did not stop the abuse categories (prior compliance). Thick and thin lines represent 90% and 95% confidence intervals, respectively. See Figure 6 in main document.*

3. Robustness

3.1 U.S. and UK Country Samples

We start with investigating whether our main results are consistent across countries. Figures below suggest that the main effects of conjoint features remain consistent except the effect of type of human rights abuse. Overall, the results grouped by each country vis à vis our hypotheses are similar to the main results with the pooled sample.

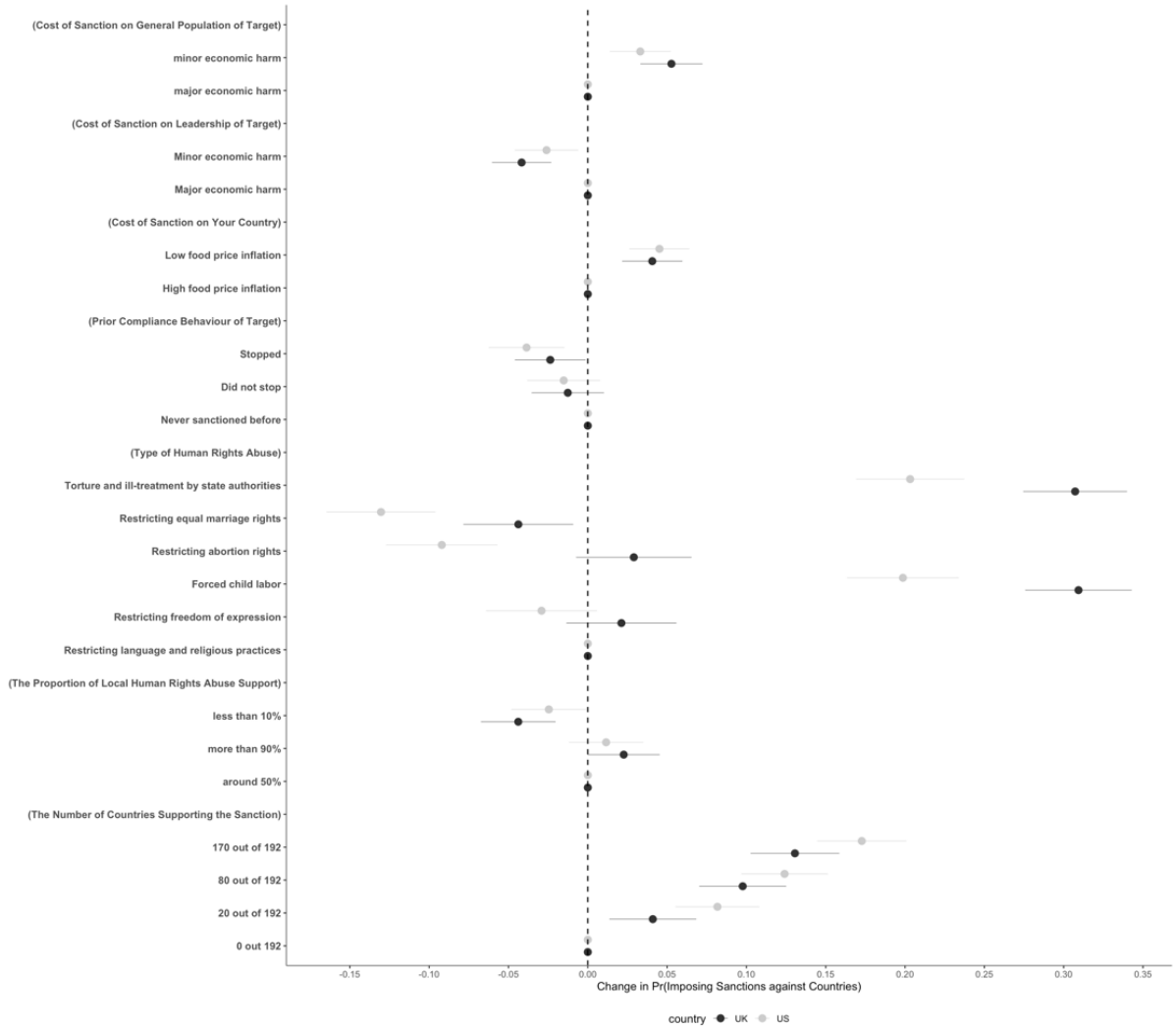


Figure 20. *AMCEs by Country*

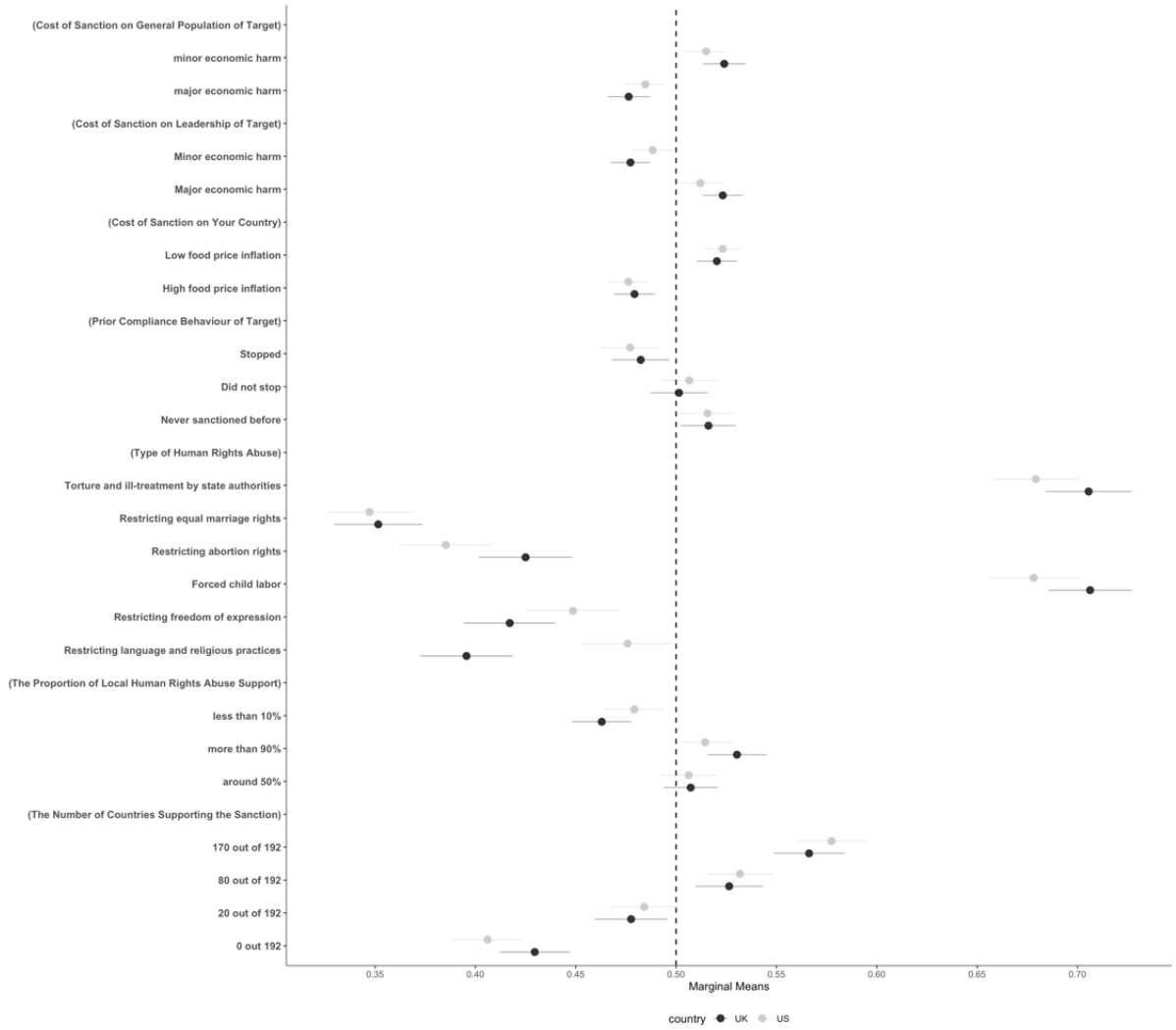


Figure 21. *Marginal Means by Country*

3.2 Rating-based Outcome Variable

We re-estimate our main results through an alternative rating-based outcome variable: "If you could vote on each of the countries in a referendum, how likely is it that you would vote in favor of or against imposing sanctions?". The conjoint features have similar effects on imposing sanctions through the rating-based outcome variable (see Figure below).

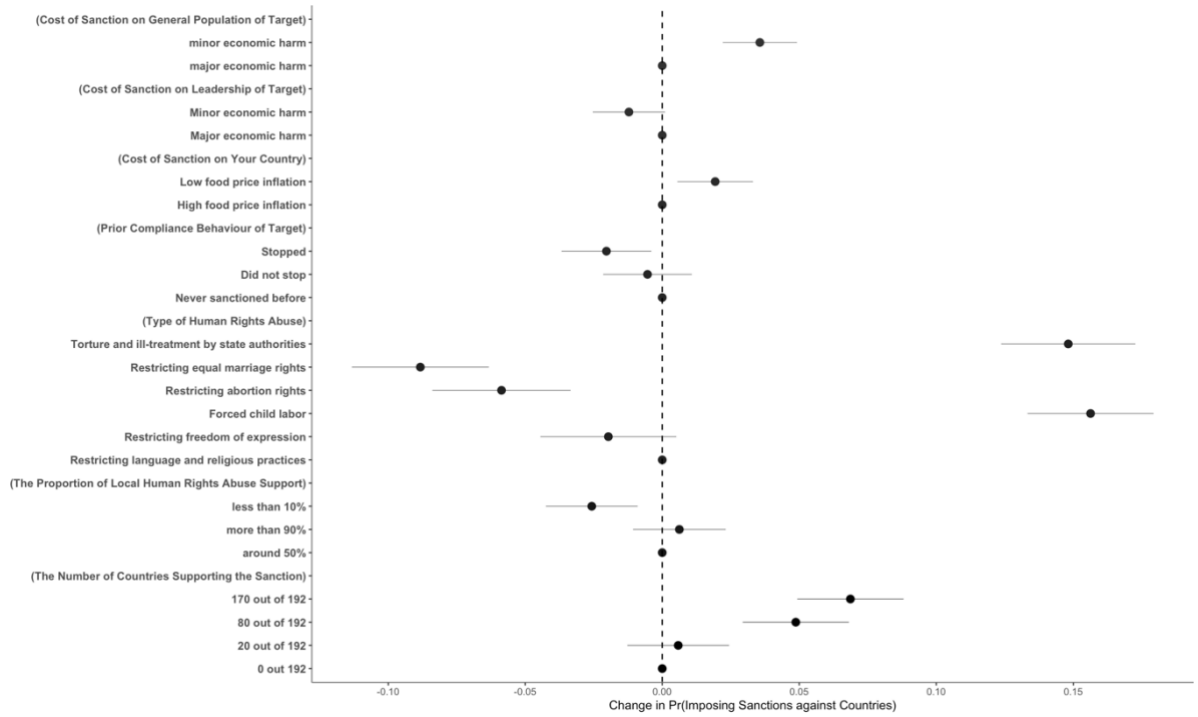


Figure 22. *Estimated Average Marginal Component Effects*

3.3 Human Rights Commitment and Behavior

We investigate how the main effects are conditioned by respondent's prior human rights commitment and behavior. Figure 23 illustrates the effect sizes for low and high human rights commitment groups. As these two respective groups follow substantively similar patterns, we conclude that our results remain robust to the level of commitment to human rights. We further breakdown this result by three human rights related past behavior: protesting against human rights violations (Figure 24), donation to a human rights organization (Figure 25), and signing a petition in support of human rights (Figure 26). Again, the results indicate similar patterns with respect to human rights related behavior.

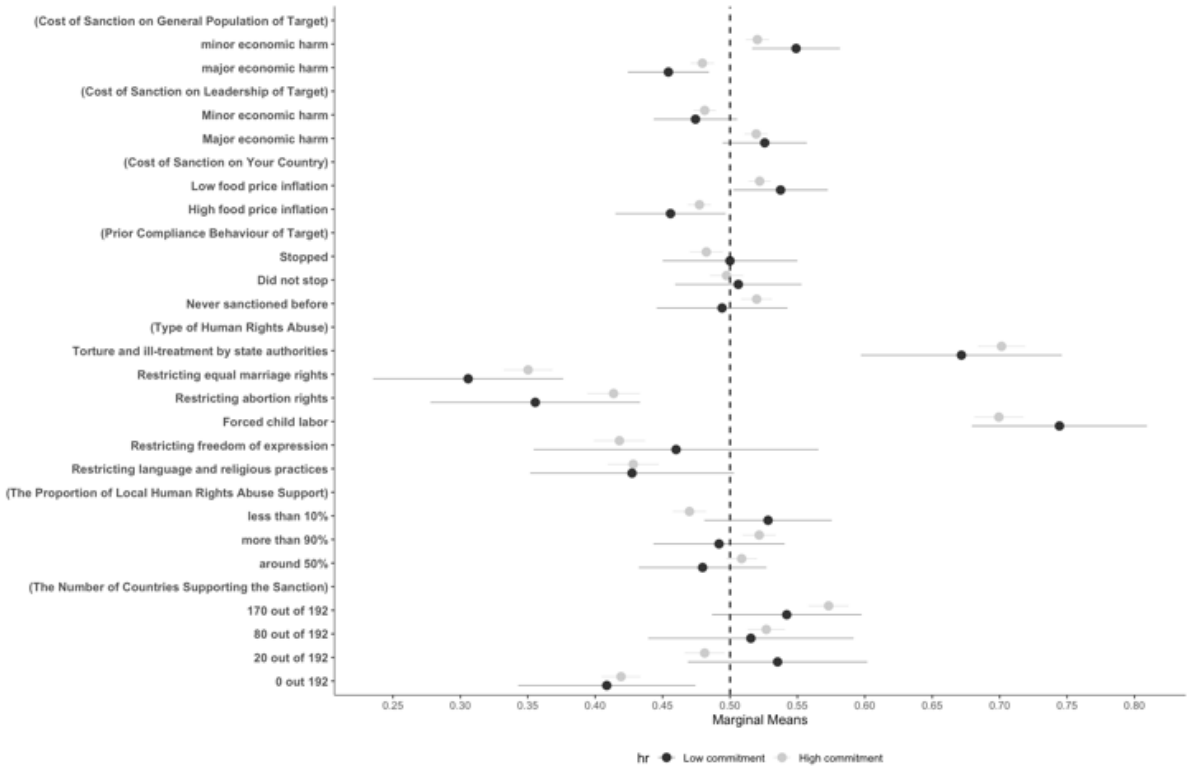


Figure 23. *Marginal Means by Human Rights Commitment*

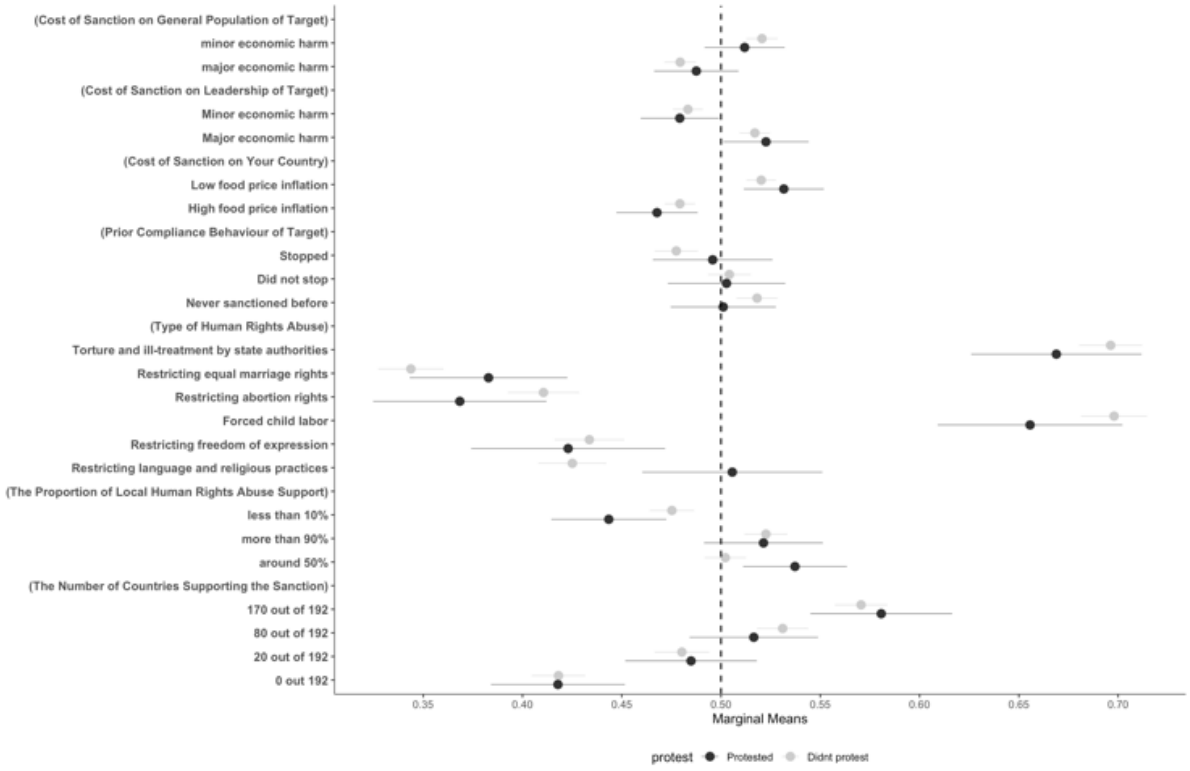


Figure 24. *Marginal Means by Human Rights Behavior: Protest*

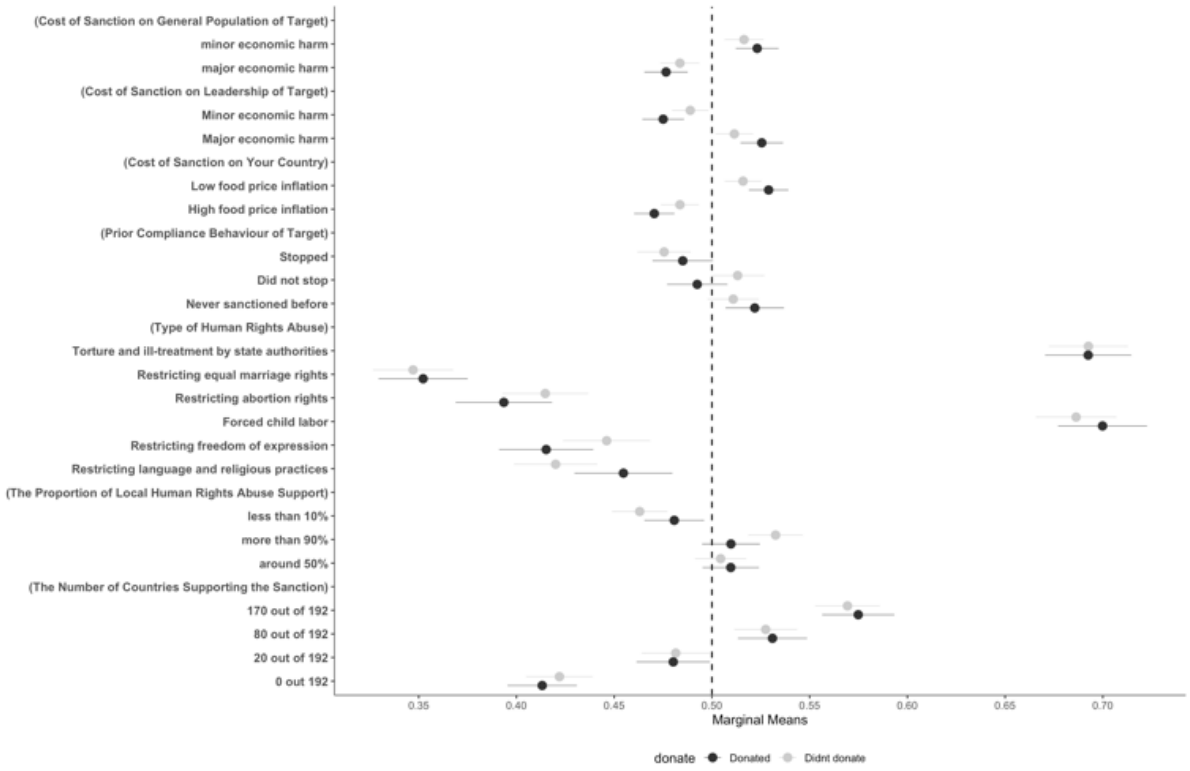


Figure 25. *Marginal Means by Human Rights Behavior: Donation*

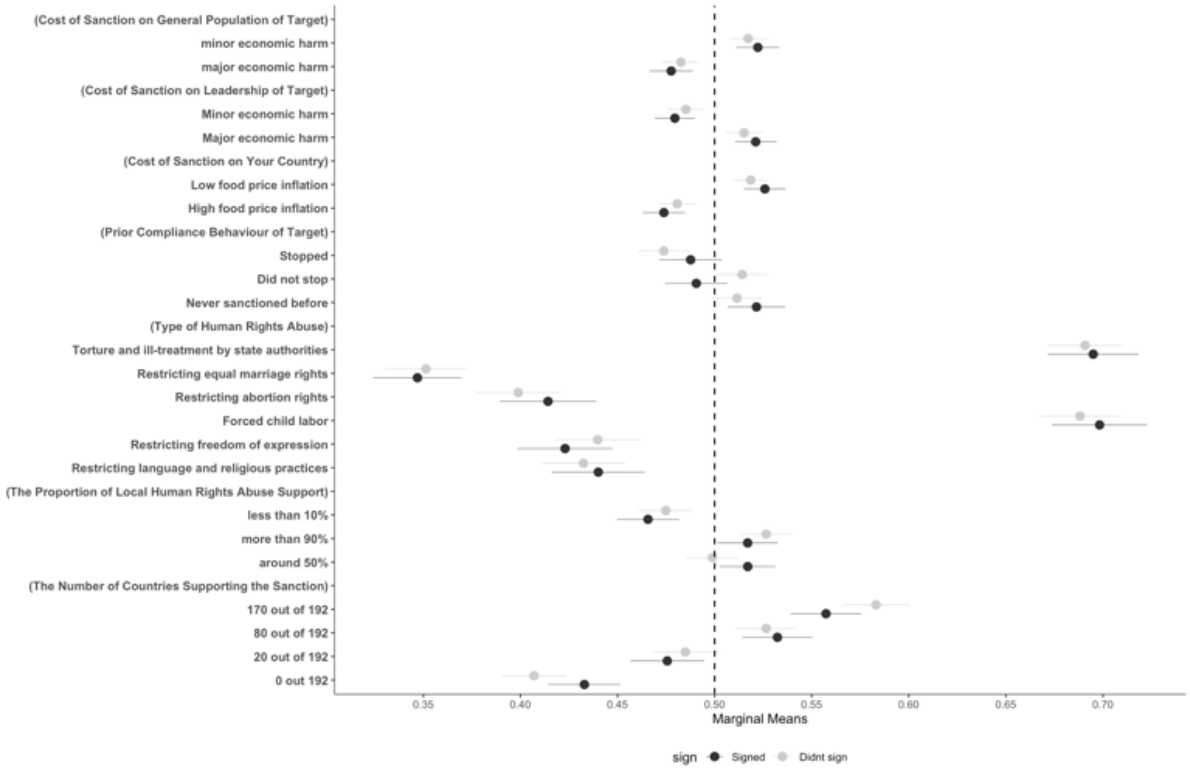


Figure 26. *Marginal Means by Human Rights Behavior: Signing a Petition*

Next, we test whether the main effects are obscured by the confidence in the influence of public opinion and the different level of perceptions towards the recognition of human rights norms in society. Figure 27 plots the main results by respondent's confidence in the influence of public opinion, while Figure 28 reports the main results by respondent's perception towards how vastly human rights norms are accepted by other citizens.

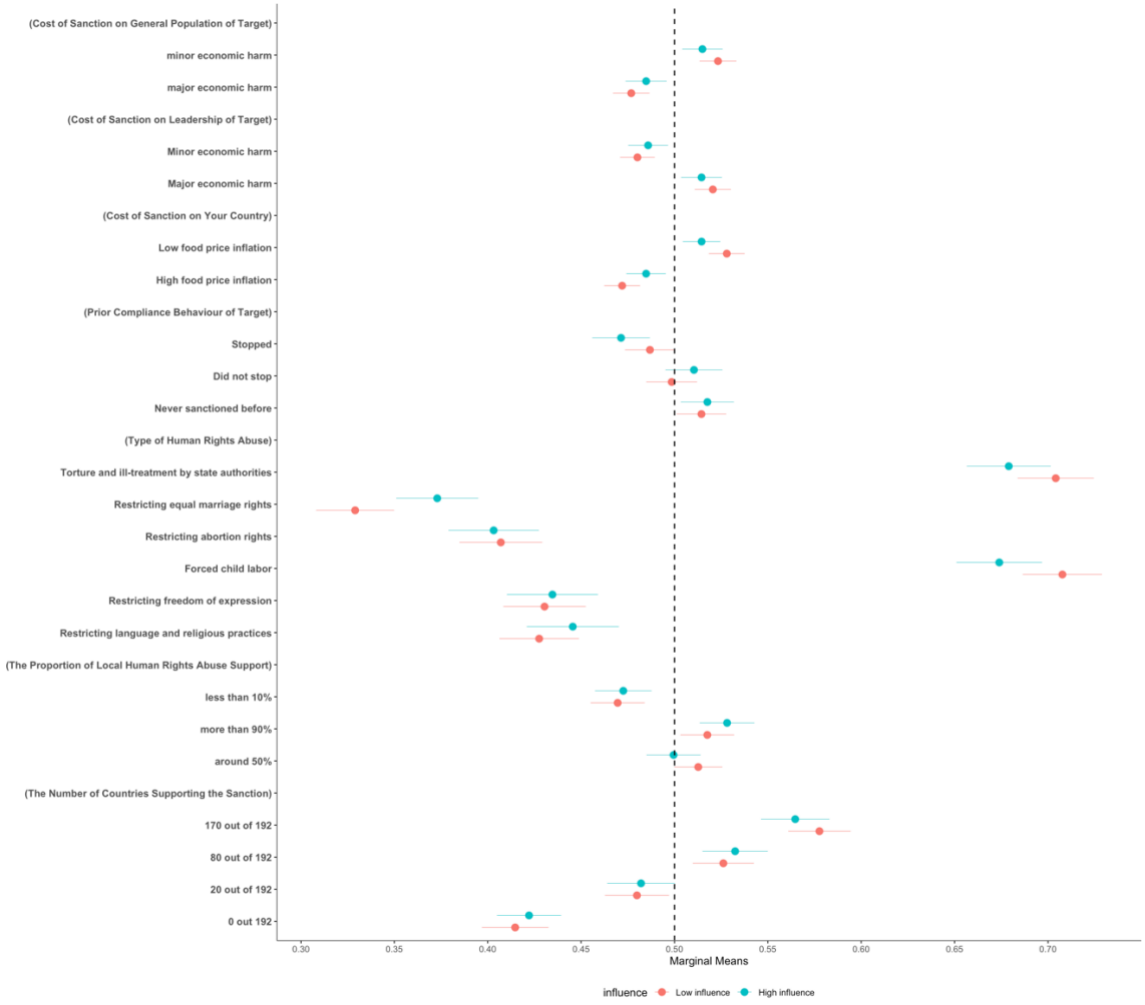


Figure 27. *Marginal Means by the Confidence in Influence of Public over Policy Decision-making*

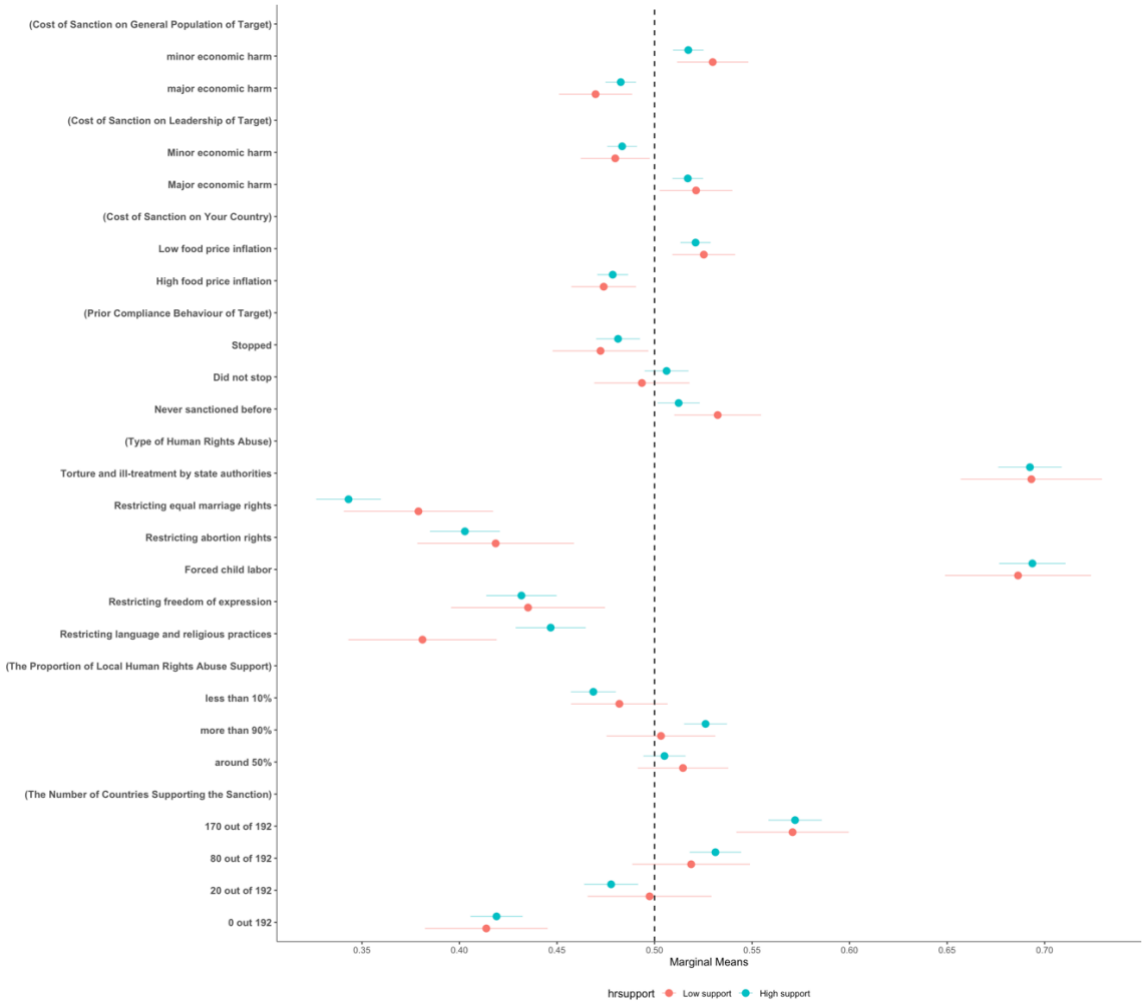


Figure 28. *Marginal Means by Perception on Human-rights Support by Others*

3.4 Carryover Effect Testing, Balance Testing, and Profile Positionality

First, we verify the assumption that respondents do not carry over the effect from one round of assessing country profiles to another round to ensure that multiple observations of each respondent can be treated as independent of one another. Figure 29 shows that this assumption holds as there is no carry over effect across rounds and attributes.

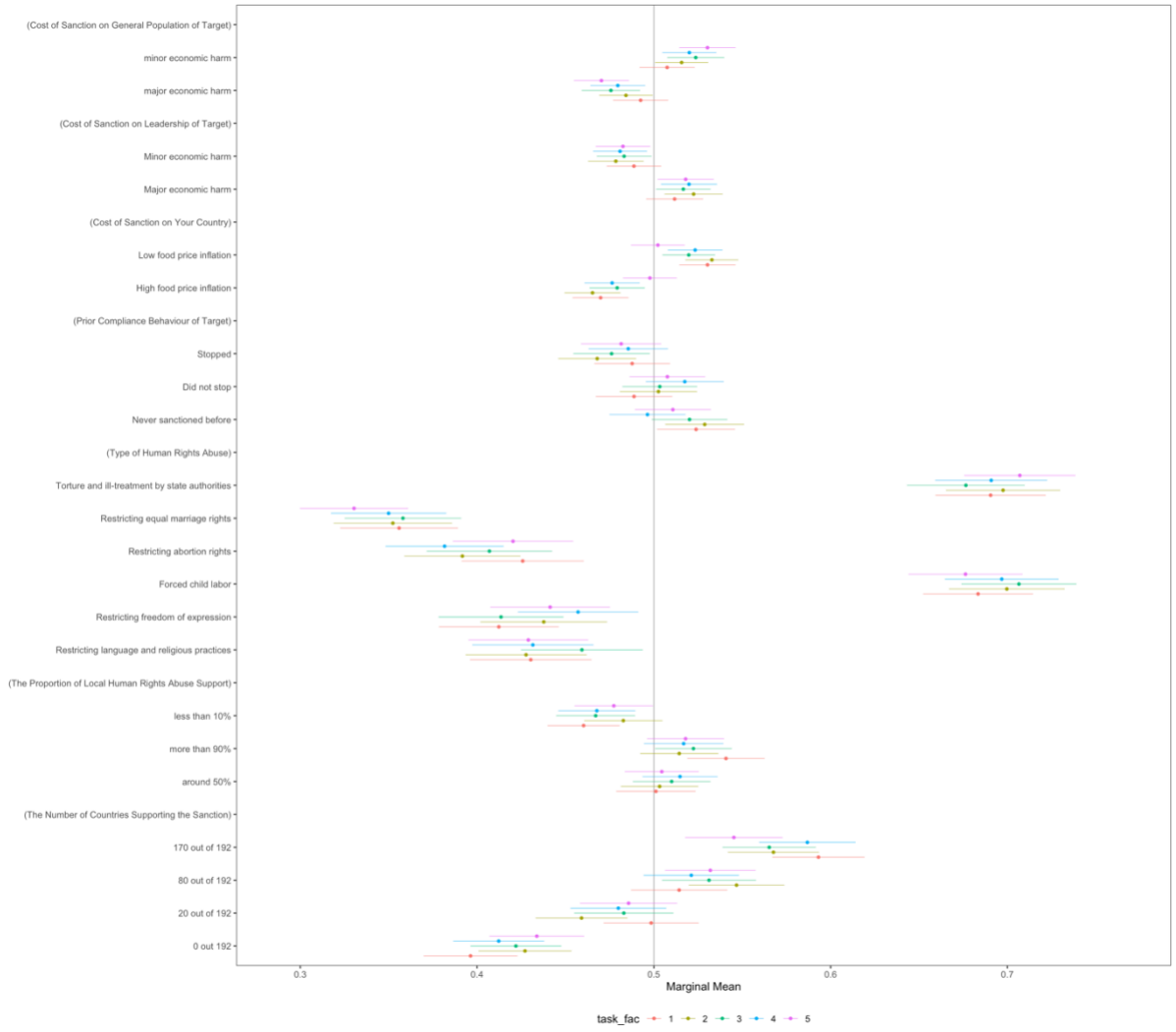


Figure 29. *Carryover Effects Testing*

Second, we plot whether the position of human-rights abusing country profiles (i.e. being listed on the left/right side of the conjoint table) affects our main estimates. Figure 30 shows that there are no substantial concerns about the positionality.

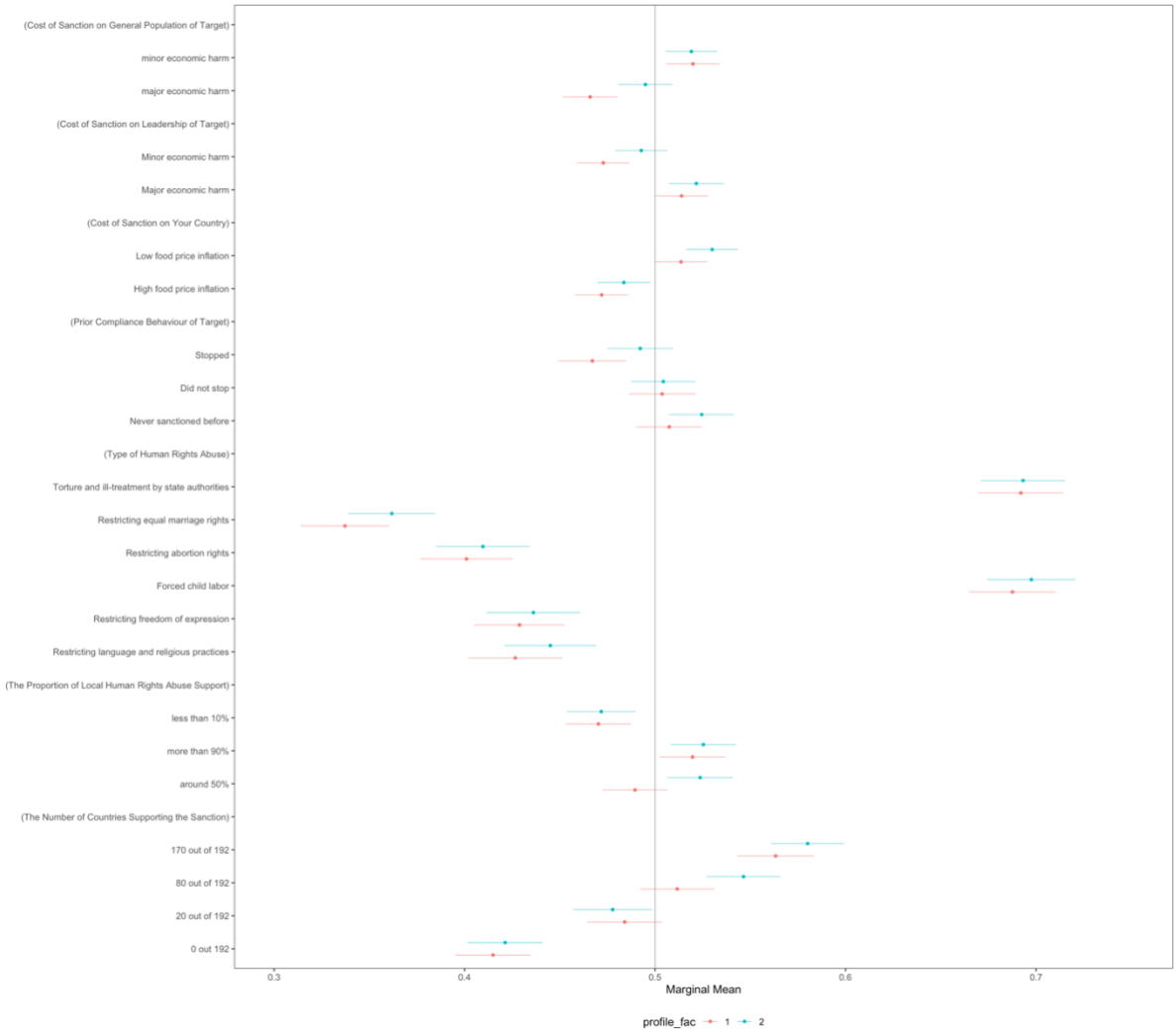


Figure 30. *Positionality Testing*

Lastly, we control whether the randomization ensures that our within-subjects design is balanced in the study. Towards this end, confidence intervals of marginal means in Figure 31 shows that each level of conjoint attributes is uniformly distributed by a covariate (respondent's age). Therefore, the potential imbalance does not significantly affect our estimates.

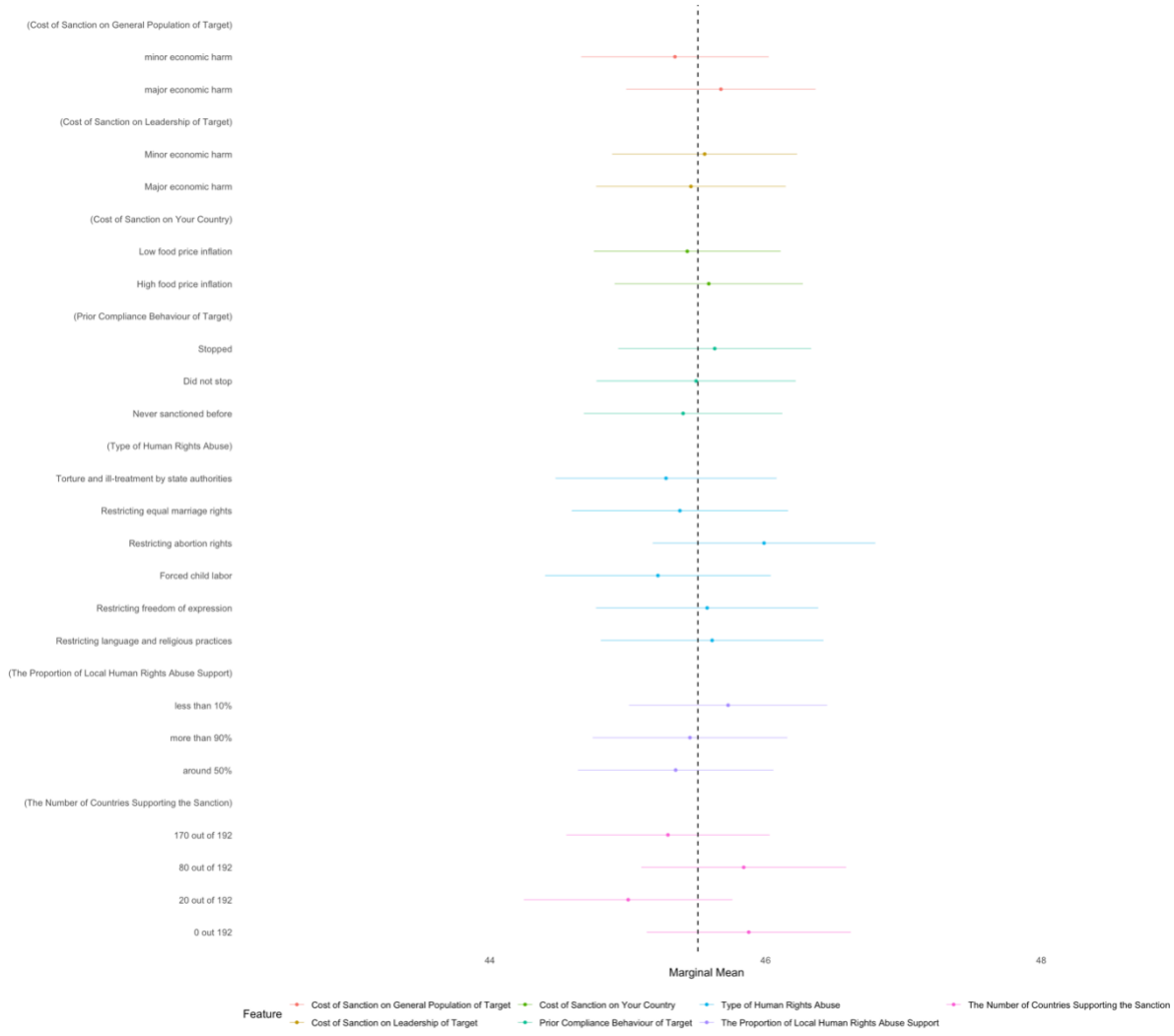


Figure 31. *Balance Testing*

References

Stefanelli, Alberto, and Martin Lukac (2020) Subjects, Trials, and Levels: Statistical Power in Conjoint Experiments.

Lukac, Martin, and Alberto Stefanelli (2020) Conjoint Experiments: Power Analysis Tool. Retrieved from https://mblukac.shinyapps.io/conjoints-power-shiny/