# Who's Afraid of Moral Trade?

Krzysztof Pelc[†]

(*Caveat lector*: early draft)

## Abstract

Social divisiveness is often seen as a barrier to interaction between individuals of opposing beliefs. This article examines a mechanism that *relies* on such differences. In a moral trade, individuals with different beliefs exchange commitments on actions pertaining to those beliefs, in a way that is mutually beneficial. If structured correctly, Democrats and Republicans, pro-life and pro-choice advocates, vegetarians and evangelists can all commit to arrangements that generate "moral gains", turning zero-sum games into positive-sum games. This article formalizes moral trade and examines some of its unique aspects, highlighting the distinction between consumption tastes and moral tastes. I suggest that the potential gains from moral trade are vast. Yet would such trades ever take place? Despite the potential for mutual benefit, moral trade may clash with prevalent norms against the commodification of principled beliefs. It is also prone to significant credibility problems: parties have reason to doubt the other side's true beliefs, and fear that they will not honor their commitments. To gauge the significance of these obstacles, I run two survey experiments on a combined sample of 4300 US respondents, strategically timed during the run-up to the 2024 Presidential election. The results indicate that large majorities appear open to moral trade, even over morally fraught issues. Individual-level variation suggests a balance of self-interested and affective factors. Moral reasoning type and religiosity both appear as significant factors. Based on the evidence, moral trade holds promise as a means of improving welfare in polarized societies, yet the design of the exchange mechanism matters.

[†] Department of Politics and International Relations, Oxford University. krzysztof.pelc@politics.ox.ac.uk.

# 1 Introduction

It is now commonplace to bemoan the rise in polarization, tribalism, and partisan divisions within industrialized democracies (Iyengar and Westwood, 2015; Reiljan, 2020; Bettarelli, Reiljan and Van Haute, 2023). Such divisiveness is faulted for a range of social ills, from a decline in trust in institutions (Brady and Kent, 2022; Oxtoby et al., 2023) to a lower propensity for pro-social behavior like blood donation and vaccination (Kim and Pelc, 2024*a*,*b*). Social sorting—the clustering of individuals into homogenous informational environments—only compounds the issue: algorithmic recommendations on digital platforms and geographic self-segregation both create online and real-world echo chambers that may amplify existing biases and contribute to affective polarization (Flaxman, Goel and Rao, 2016; Duskin et al., 2024; Liu, Andris and Desmarais, 2019). The result is that individuals are not only less willing to hire, buy from, work for, or date members of an opposing ideological camp (Iyengar, Sood and Lelkes, 2012; Huber and Malhotra, 2017); they are less likely to encounter them at all.

While conventional wisdom thus views more divisive beliefs as barrier to interaction between individuals of opposing beliefs, this article examines a mechanism that *relies* on the existence of such differences, and holds the potential of extracting welfare improvements from them. Moral trade involves actors with differing beliefs exchanging commitments on actions pertaining to those beliefs, in a way that is mutually beneficial. If structured correctly, Democrats and Republicans, pro-life and pro-choice advocates, vegetarians and evangelists can all commit to arrangements that can generate "moral gains", turning zero-sum games into positive-sum games. The idea was introduced in Ord (2015), who highlighted its parallels with traditional commodity trade. Just as gains from commodity trade are driven by differences in tastes, so too are the gains from moral trade driven by differences in "moral tastes," or principled beliefs. In both cases, it is the existence of heterogeneity that allows for mutually beneficial exchanges. In a moral trade, the "gains" to Self from an exchange with Other take the form of a closer alignment of Other's behavior with Self's beliefs than would otherwise obtain. If an arrangement is reached where both sides are morally better off by exchanging such commitments, then a moral trade is said to take place.

In this article, I extend the foundation laid by Ord (2015), exploring not only the parallels between commodity trade and moral trade, but also their key distinctions. At the heart of this comparison is the nature of preferences in each domain. Unlike commodity trade, where preferences tend to be self-contained, moral preferences are inherently concerned with *other actors'* preferences: an individual's utility diminishes as the gap between their own moral preferences and those of others widens. The promise of moral trade is that it can help bridge that gap; as I show, that is also its limitation. This distinction between the nature of preferences in the two realms may be the best criterion for delineating the otherwise blurry contours of moral trade versus commodity trade.

Building on this theoretical distinction, I examine how commodity trade and moral trade intersect. Modern trade agreements, for example, often include provisions for labor, gender, or human rights standards—provisions that have been decried from a strict economic perspective as cynical tools for import-competing groups to erode the comparative advantage of poorer nations. Yet this instrumental view appears insufficient, as at least some of these provisions reflect genuine "collective preferences" (Charnovitz, 2005). The lens of moral trade allows us a novel perspective on such provisions: these can be recast as mutually beneficial moral trades embedded within commodity trade agreements, where richer countries provide market access in exchange for a closer global alignment with their populations' collective beliefs. This framing forces us to re-examine the relationship between country wealth and the scope of moral concerns, whereby a country's moral universe may expand beyond its national boundaries as its level of development rises.

I then shift to an central empirical question: given the tremendous gains to be had, would moral trades ever take place? There remain considerable challenges, including objections rooted in moral principles themselves, credibility concerns, as well as affective factors. I weigh the importance of these challenges through two survey experiments conducted on a combined sample of 4300 US respondents. I also use these to examine whether completing a moral exchange, by itself, might lower animosity to the opposite ideological camp. The findings shed light on the structural factors and individual traits that drive attitudes to moral trade, offering insights into how such exchanges might succeed in divided societies.

Given the novelty of the concept, some illustrative examples will be helpful. In the first case, consider an encounter between a vegetarian and an advocate for open-source software. The first is morally averse to eating animals, the second is just as opposed to proprietary software. If the open-source advocate has weak preferences over his diet, and the vegetarian feels similarly indifferent about the software she uses, they can reach a mutually beneficial arrangement. The vegetarian commits to eschewing proprietary software for a year, and in exchange, the open-source advocate refrains from meat over the same period. As I go on to demonstrate, moral "pet issues" of this type, where individuals hold intense preferences on a given belief parameter and identify with it to the exclusion of all others, increase the opportunities for moral exchange, and thus moral gains. The mechanism is akin to issue linkage in bargaining theory. Self's relative *indifference* on Other's moral issue of interest represents their "productive potential", as it defines the commitments Self can make at low-cost to fulfill Other's preferences. Yet orthogonal moral issues become harder to find under political polarization, as political, cultural, religious, and moral beliefs cluster on partisan identities. Cross-cutting identities are what moral trade relies on; these grow rare as a result of polarization.

While the values being traded on in a moral trade can be orthogonal to one another, exchange is also possible, albeit more contentious, when two agents hold different beliefs on the *same* ideological spectrum. This is the type of trade I focus on in the empirical section, precisely because it appears as a hardest-case test. In the highly polarized two-party setting of US politics, a Democrat and a Republican intent on donating to their respective causes might each commit to refraining from doing so, and rechannel their money to a non-political third cause instead, such as an animal shelter or public park conservation. If each moral agent believes that their donations would have cancelled out the other—neutralizing any net effect on the dimension of partisan politics—and if they assign some positive value to an orthogonal cause, the arrangement generates net moral benefits from each agent's standpoint. The non-political cause gets funding, while the political causes are no worse off than they would have been.[1] Here, too, the "trade space" decreases as all issues are subsumed by the partisan dimension, and orthogonal issues become harder to find. Similar arrangements might be struck in the realm of reproductive rights, between pro-life and pro-choice advocates keen to donate to their respective camps. In the empirical section, I examine individual preferences over moral trades on both these politically fraught issues. The aim is to show that even deeply divided individuals can plausibly find common ground by structuring their commitments to maximize mutual moral gains.

In a third instance, mutually beneficial exchanges can also obtain when both sides agree on the moral value to be maximized, but hold this preference at different intensities, akin to differences in marginal rates of substitution in conventional trade. Consider the archetypal debt-for-nature swap organized by Conservation International in 1987. The NGO purchased $650,000 of Bolivian debt at a discount in return for Bolivia's establishment of protected areas in the Amazon. That deal was the prototype for the more than 140 debt-for-nature swaps that have since taken place.[2] A similar logic underpins the Amazon Fund, a contemporary initiative supported by wealthy nations like Norway to reduce deforestation in the Amazon, by compensating Brazilian farmers who stand to gain from turning the forest into arable land (Birdsall, Savedoff and Seymour, 2014). Donor

[1] Below, I examine the assumptions this second claim rests on in greater detail.

[2] "How one South American country became a lab for conservation". Oct 26, 2023. Conservation International.

countries act on their moral convictions about climate change and biodiversity. Brazilian farmers, for their part, may also value the preservation of the Amazon on normative grounds, but assign a lower *relative* weight to this belief, due to different resource endowments. As Norway and Brazil thus have different marginal rates of substitution for rainforest preservation, a trade becomes possible: Norway commits development funds to Brazilian farmers, who in exchange commit to not developing the rainforest for agriculture. As opposed to traditional externalities that are evaluated from a societal standpoint, the transfer leads to a Pareto improvement by internalizing a moral externality from the standpoint of a specific agent—in this case, Norway. Such international arrangements push towards what may be thought of as "moral price equalization," akin to factor price equalization in the Heckscher-Ohlin model. The trade continues until the relative valuation of rainforest preservation converges across parties.

What are the main obstacles to moral trade? In the remainder of this article, I test some of the empirical implications of a model of moral trade. Indeed, the gains to be had from moral trade on a large scale are potentially vast, yet these face serious obstacles. First, individuals may hold moral reservations about moral trade itself, no matter the gains to be reaped, if moral trade also runs afoul of prevalent norms against commodification of principled beliefs, along the lines of "taboo tradeoffs" (Tetlock et al., 2000). Secondly, moral trade is prone to a significant credibility problem: as I demonstrate, trading parties have reason to doubt whether whether the other side has correctly represented their beliefs, and whether they will abide by their end of the bargain, since there is an incentive to misrepresent on both counts. Given these competing considerations, would moral trade ever take place? To find out, I run two studies on a combined sample of over 4300 US respondents, strategically timed during the run-up to the 2024 Presidential election. The survey offers four main empirical findings.

First, large majorities appear in support of moral trade. Yet some issues are associated with greater support than others: individuals are thus significantly more open to trades over political partisanship than equivalent deals over reproductive rights, though both see majority support. Second, across both issues, individuals with stronger views are significantly *less* likely to agree to a deal, in ways that comport with formal expectations. Yet counter to formal expectations about "moral gains", increasing the counter-party's level of ideological commitment does *not* result in greater willingness to trade. This may speak to a countervailing reluctance to transact with more committed moral opponents. While credibility emerges as the dominant practical concern—particularly among those with strong views—the second dominant concern is discomfort with the commodification of beliefs. Third, and surprisingly, an attempt to normalize moral trade by discussing vote-trading among US legislators makes individuals significantly *less* open. One explanation is that individuals in fact disapprove of such deals at the political level, and that disapproval spills over onto their individual decisions. Fourth, a number of individual traits appear predictive of attitudes towards moral trade. I measure the extent to which individuals rely on consequentialist versus deontological reasoning, and find that consequentialists are significantly more open to moral trades, regardless of the issue. Conservatives appear significantly less open to moral trade than Liberals across all issues. And religiosity matters: atheists are considerably more open, and even *among* believers, religious salience is negatively related to openness to moral trade to a significant degree. Yet given the average levels of support, moral trade appears to hold promise as a welfare-enhancing mechanism in polarized societies. The findings hold implications for how a moral exchange system ought to be designed to maximize mutual gains. At its best, moral trade offers a framework for harnessing value differences to achieve collective gains, by identifying neutral or overlapping moral beliefs.

# 2 Theory

## 2.1 The Scope of Moral Trade

Trade allows individuals to consume more of what they value, generating mutual utility gains through exchange. This is one of the most fundamental insights of economic theory: trade expands welfare by reallocating resources to better align with individual preferences. Trade opportunities are typically said to arise from differences in initial endowments or productivity, as per the theory of comparative advantage. Yet another driver of trade comes from differences in *taste*—the subjective preferences that determine the extent to which individuals value specific goods and services. If one person likes raspberries while another prefers blueberries, they can conduct a berry trade, allowing each party to better align their consumption bundle with their preferences. Even if both individuals are equally adept at procuring raspberries and blueberries, trade enables both to consume more of what they value, increasing overall welfare. In this way, utility does not derive only from increasing total consumption, but also from better tailoring consumption to individual preferences. Trade thus serves as a mechanism to expand the realm of possibility, transforming initial constraints into opportunities for mutual gain.

The concept of moral trade extends these principles to the domain of moral values. Individual moral utility can be thought of as the extent to which the world aligns with one's moral preferences. Gains in utility then represent changes in the realization of those preferences. In moral trade, each agent gains from outcomes that better align with their moral commitments.[3] This minimalist definition has the benefit of applying across various moral frameworks. Moral trade then maximizes utility in much the same way as commodity trade, but here utility is derived from satisfying ethical beliefs. Just as commodity trade leverages differences in preferences to create mutual benefit, moral trade leverages differences in moral beliefs. This is the main insight in Ord (2015)'s original treatment, where he insists on the commonalities between moral trade and commodity trade. But it leaves open the question, what exactly distinguishes moral trade from conventional trade? Indeed, conventional trade also routinely involves exchanges of morally significant goods or services with moral effects, between market actors who value these differently. Think of trade in sustainably sourced goods or fair-trade products, or goods produced by prison labor. Are these instances of moral trade? How would we know?

The key distinction between the two settings, I argue, lies in the relation between agents' preferences. In particular, moral utility implies that Self holds preferences over Other's beliefs and the actions pertaining to those beliefs. Each agent gains or suffers by the degree to which their moral beliefs are aligned with the other agent's. They would thus be best off if everyone else held identical beliefs to their own. In the case of conventional trade and consumption, one might argue the exact opposite. When picking between possible worlds, individuals might prefer to live in a setting where consumption tastes are diverse. If one could "desire by pill" Millgram (1997), there is reason to think that we would be better off changing our consumption preferences to be different from those of others.[4] This not only because preference heterogeneity can reduce competition and alleviate scarcity effects, but precisely because diverse preferences expand the trade space, in ways that allow each individual to consume more of what they value. In the example above, the market actor who likes blueberries gets more of what they value *because* the other prefers raspberries. If they both held identical berry preferences, they would both get less of what they value.

[3] This is by no means the only possible conceptualization of moral utility. An alternative would focus on moral agency, and see utility in the extent to which an agent's actions have shifted the world toward their moral ideal. I believe this would not affect choices in moral trade, since the aim of maximizing gains from moral trade would remain; yet it would have implications for overall utility calculations, since it would affect the baseline levels of moral utility absent trade.

[4] This claim rests on assumptions about intra-personal well-being comparisons. Individuals need to be able to meaningfully compare their utility across different sets of preferences, an area of considerable ambiguity in orthodox economics. Yet this seems at least plausible under bounded rationality assumption, where agents can identify opportunities to enhance their well-being through preference adaptation. Since this discussion goes beyond the scope of the paper, I set it aside for now.

By contrast, in the moral trade setting, each individual would be best off if everyone else shared their beliefs, *even if* there is a well-established frictionless mechanism for moral trade, and all possible moral trades are consummated. That is, trade can only ever be a means of bridging the gap. Moral agents may agree to and gain from moral trade, yet they would still elect to live in a world where others' moral values were so aligned with their own as to void the possibility of any trade. To be sure, preferences in the traditional economic realm can also have an interdependent aspect. The value of "network goods" increases as more people want and consume them. "Bandwagon goods" similarly grow more valuable as more people desire them. In both these cases, one market actor's preferences depend on those of others, but this dependence is pragmatic and instrumental: shared preferences create utility by enhancing coordination, visibility, or social signaling. By contrast, in the moral setting, the interdependence between preferences is inherently normative and prescriptive. The reason for seeking alignment is that moral agents perceive others' divergent beliefs as not just different, but wrong. Behavior based on those moral beliefs is a manifestation of this misalignment. By reducing misalignment, moral trade can transform clashing beliefs into cooperative outcomes, benefiting all parties by allowing a closer alignment of action with shared values.

This aspect of preference interdependence is present in all the examples cited in Ord (2015). It is also present in my three introductory examples above, where the two parties to an exchange (i) hold moral beliefs on two distinct parameters; (ii) occupy opposite positions on the same parameter; and (iii) share the same value on the same parameter, but at different intensities. In each case, each party would prefer for the other to have identical beliefs to their own.[5]

It is helpful to consider what types of trade this criterion of preference interdependence excludes. Consider two work colleagues, one Jewish and the other Christian, who place moral importance on celebrating different holy days. They agree to trade work shifts: the Jewish colleague works during Christmas, while the Christian colleague works on Shabbat. This arrangement benefits both parties and increases their moral utility, as each is able to honor their sacred days. However, these gains do not result from preferences over the other's moral behavior. The Jewish

[5] This may not appear to be the case for Brazil in the context of the Amazon Fund: it seems as if Brazil benefits from developed countries' stronger relative valuation of the Amazon, since this difference is the source of Norwegian development funds. Yet this depends on what Brazil puts on the other side of the scale. To illustrate, if Norwegian funds are directed towards economic development, then imagine that Brazil is trading off preservation of the Amazon against its moral valuation of poverty relief. If so, then the trade can be said to arise from Norway's higher relative weighing of the Amazon's preservation, compared to poverty relief in Brazil. If so, then Brazil would indeed prefer that Norway's moral beliefs were more aligned with its own, such that Norway would fund poverty relief in Brazil and rainforest preservation at the same ratio as Brazil itself.

colleague does not seek to influence the Christian colleague's moral practices, and vice versa. The trade works precisely because their beliefs are limited to their own actions—they are parochial rather than universalist. If given the choice, both individuals would prefer a world with preference diversity, as it creates the conditions for this mutually beneficial exchange. Although the exchange involves moral preferences and produces moral utility, it therefore does not qualify as a moral trade under the criterion outlined above. It is, instead, a conventional trade: the traded services themselves may be infused with (parochial) moral value, but the scenario lacks the preference interdependence that defines moral trade.

## 2.2 Contemporary Trade Agreements as Moral Trades

Consider an analogous example at the international level. It is often claimed that from an economic standpoint, a developing country's lower labor standards represent its comparative advantage vis-à-vis developed countries. This is not to say that developing countries do not value high labor standards; rather, countries at a lower level of economic development are willing to trade off higher labor standards in exchange of economic development, in ways that allow them to produce goods more cheaply. In this sense, Brazil might be said to "specialize" in producing lower-labor-standards goods, exporting these goods to wealthier countries like Norway, where labor standards are higher, and domestic labor is therefore more expensive. Norway effectively trades its preference for higher labor standards by importing goods produced at lower standards, benefiting from their lower cost. Here, too, commodity trade is made possible by moral parochialism: Norway does not value labor rights writ large; it values its own labor rights. This is a conventional trade scenario: just as with the Jewish and Christian work colleagues, as long as trade is possible, both actors can be said to be better off by virtue of holding different preferences.

Then, if the scope of Norway's moral values expands, and these become less parochial and more universalist—as often happens with a rising level of development—this will interfere with regular commodity trade, insofar as Norwegian consumers may object to products associated with lower labor standards. The expansion of the scope of Norwegian morality reduces Brazil's conventional comparative advantage. Yet such an expansion of moral scope is precisely what makes moral trade possible. As Norway's moral beliefs extend to Brazil's workers, it may enter into a genuine moral trade, along the lines of the Amazon Fund: Norway might offer e.g. greater market access in exchange for greater alignment by Brazil with Norwegian moral values over labor standards.

This shift from commodity trade to moral trade is arguably what many modern trade agreements amount to. As Rodrik (2018) has observed, many contemporary trade agreements do not appear to be primarily about lowering trade barriers at all. Instead, they are largely concerned with regulation, standards, and other "behind-the-border" issues that have ambiguous economic effects. In particular, industrialized democracies increasingly negotiate legal provisions that are meant to get other countries to align with their regulatory priorities. Pascal Lamy, the former Director General of the WTO, used the term "collective preferences" to describe the genuine differences in social priorities between countries.[6]

Lamy distinguished between collective moral preferences that have a universal aspect, pertaining to the behavior of market actors beyond the nation, and those that have a strictly parochial aspect, limited by the country's borders. As he observes, and parallel to the example of the work colleagues above, "[t]he Jews and the Lombards of the Middle Ages were able to operate only because Christianity did not demand of 'foreigners' what it demanded of believers." But as Lamy goes on to note, there are also "fundamental social standards" that "are considered by their promoters to be universal and applicable to all. This sets limits to the working of comparative advantage." My point is that while they do impose limits on (conventional) comparative advantage in traditional trade, the existence of such collective preferences may offer countervailing opportunities for moral trade.

Collective preferences of this type are often portrayed as ploys by import-competing industries in wealthy countries to reduce the comparative advantage of developing countries (Charnovitz, 1991); yet in a growing number of cases, as with the EU's due diligence law to enforce environmental and human rights abuses in companies' supply chains, its deforestation initiative, its new regulation on plastic packaging recycling, and arguably even its carbon border adjustment mechanism (CBAM), it seems we are dealing with an attempt at moral alignment, rather than commercial interest.[7]

Today's "deep" trade agreements thus represent a particular kind of arrangement. The inclusion of ambitious environmental, labor, and human rights standards are not so much constraints on trade that come at the cost of efficiency, but the pursuit of different type of gain through a different mechanism: moral trade. These provisions represent a large-scale moral trade within a commodity trade agreement. The purpose is not so much to increase and shape trade flows, as to

[6] As Lamy defined them, "Collective preferences are the end result of choices made by human communities that apply to the community as a whole... In short, they should be seen as values." Pascal Lamy EU Trade Commissioner The Emergence of collective preferences in international trade: implications for regulating globalisation. Brussels, 15 September 2004.

[7] See, e.g. Beattie, Alan. March 21, 2024. Financial Times. "The global downside of European consumers' green principles."

align Foreign's behavior with Home's moral preferences.

There is a likely trade-off between the two realms: as the benefits of commodity trade expand—through cheaper and more plentiful goods—their incremental value diminishes, making moral gains from trade agreements relatively more attractive. At some point, societies may willingly sacrifice additional material benefits from conventional trade to achieve greater moral gains, as represented by greater alignment of Foreign's behavior with Home's moral values. Viewed through this lens, trade agreements become instruments for optimizing not only economic efficiency through an increased trade volume, but also moral efficiency. The total gains from trade agreements should thus be understood as the sum of material and moral gains, with moral trade enhancing the value of trade agreements in a way that complements traditional economic objectives. The apparent overloading of trade agreements with so-called "non-trade" issues may in fact represent an optimal trade agreement, with "trade" understood in its dual sense.

## 2.3 Economic Parallels

**Consumption versus Production** The distinction between production and consumption is unambiguous in conventional trade, but it becomes more difficult to make out in the moral setting, where both follow from the parties' beliefs. Here, Self's production is the enactment of moral value from Other's standpoint; Self's consumption is the realization of her moral preferences from Other's better alignment with them. The distinction between consumption and production may thus seem blurred, but it is a distinction worth preserving.

In conventional trade, relative production capacity is typically determined by objective factors like technology, labor, and resource availability. By contrast, the relevant production capacity for moral value, because of how Self's utility proceeds from Other's revealed preferences, is jointly determined by the moral beliefs of each party. Looking at consumption value, in conventional trade it typically derives from objective properties of goods like functionality and scarcity. In moral trade, value is entirely subjective and relational, depending on how actions align with each party's moral framework. A trade is possible as long as the production cost is less than the consumption value; that is, as long as the cost of a moral commitment on a given dimension by one party is less than the value of that greater alignment for the counterparty.

In both settings, production capacity is determined by cost ratios, and consumption capacity is determined by the marginal rate of substitution (MRS). In standard trade theory, an equilibrium is reached when the marginal rate of substitution—the rate at which a consumer is willing to trade one good for another—equals the cost ratio, ensuring efficient allocation, and no further possible gains from trade. The same applies in moral trade. An equilibrium emerges when the cost ratio of producing moral value for the other party aligns with the willingness to substitute moral preferences for those of that other party, as reflected in the MRS. In equilibrium, the vegetarian's and open software advocate's MRS equalize, matching the moral cost ratio, and ensuring no further moral utility can be gained through additional adjustments. This means that the effort or sacrifice one party incurs to act in accordance with the other's values matches the other's willingness to reciprocate.

In this case, the possibility of a trade between the vegetarian and the open-source software advocate comes from how the vegetarian has few moral commitments against open-sourced software, and the software advocate similarly has few moral commitments against vegetarianism. That is, each party feels strongly about their issue, and weakly about the other's. The vegetarian's trade-favorable cost ratio is determined by her indifference to the software issue, compared to how highly her counterpart values it, and vice-versa. Each moral agent can produce moral value efficiently for their counterpart by acting on their own relative indifference, while the other party "consumes" the outcome in alignment with their strongest moral conviction. The existence of such moral 'pet issues', where individuals intensely value one moral dimension while remaining relative indifferent to others, creates opportunities for moral exchange that generate mutual moral gains.

**Comparative Advantage** The above reasoning also allows us to transpose comparative advantage, the most fundamental concept in trade theory, onto the moral setting. Just as countries benefit from specializing in goods where they have a lower opportunity cost, individuals in moral trade have an incentive to act on values over which they are relatively indifferent, but that are highly valued by others. Thus, the vegetarian's comparative advantage lies in adopting open-source software, which for them incurs a negligible moral cost, while the open-source advocate's comparative advantage lies in adopting vegetarianism, as it does not conflict with their moral beliefs. By leveraging these differences in moral cost structures, moral trade enables both parties to achieve gains that would be impossible through unilateral action. In the case of Norway and the Amazon Fund, Brazil's comparative advantage lies in its relative higher opportunity cost of preservation. The trade arises because Norway's willingness to pay to preserve the rainforest exceeds Brazil's economic gains from deforestation.

**Moral Price Equalization** Until what point will a moral trade like the one between Brazil and Norway remain favorable? As long as the cost ratios differ. The Amazon Fund thus functions as a mechanism for what we might call "moral price equalization." Recall that the factor price equalization theorem, associated with with Samuelson (1948) and the Heckscher-Ohlin model, posits that under free trade, the prices of factors of production, such as wages and land rents, should converge across countries even if the factors themselves cannot move freely between countries. Trade allows countries to specialize according to their comparative advantage, equalizing factor prices through market competition, rather than factor mobility. In this case, the Amazon Fund enables parties with differing moral vs economic valuations to reach a Pareto-improving arrangement, much like a market facilitates gains from trade based on differing cost structures.

Examining such arrangements through the lens of moral trade forces us to reconsider the relation between wealth and morality. Money can be used to further moral goals, as with donations to moral causes. By symmetry, one may also be forced to trade off money and morality. It follows that poorer individuals, or countries, will face a different tradeoff between money and moral values than richer ones. Brazil may thus attribute positive value to the rainforest, but its level of economic development means that at the margin, it may be willing to trade off more rainforest preservation in exchange for development than a developed country. It has a lower preference intensity on this moral dimension, compared to its opportunity cost of economic development.

One expectation is that as wealth levels converge, so will each party's marginal moral rate of substitution. As basic material needs are met, individuals shift their priorities to increasingly post-materialist concerns such as environmental sustainability, human rights (Inglehart, 2013). Consumer behavior seems consistent with this expectation, whereby higher income individuals are more willing to pay a premium for goods associated with higher environmental standards and labor rights. This is in line with Lamy's concept of collective preferences. As he puts it, "The very formation of collective preferences is dependent on income levels: the trade-off between greater affluence and environmental protection or between greater affluence and reduction of inequalities changes with income levels." In this way, moral trades such as embodied by the Amazon Fund achieve convergence of moral prices from both sides, by transferring both capital funds and moral gains where they are most valued.

## 2.4 The Social Limits of Moral Trade

This article's motivation lies in our uniquely polarized moment. From the standpoint of moral trade, polarization has two distinct relevant aspects. It represents more spread on a given moral dimension, like partisanship. But it also entails an overall collapse in the number of independent dimensions. Both of these have implications for the opportunities for moral trade.

Political partisanship is no longer merely a set of policy preferences, but functions as a deep-seated identity (Green, Palmquist and Schickler, 2008). As a result, polarization is no longer just about policy disagreements, but increasingly fuses with cultural, social, as well as moral identities. This clustering of identities aligns once-separate parameters—religion, ethnicity, and geography—under a single overarching dimension: partisanship. Technology likely has a compounding effect. Recommendation algorithms amplify engagement by offering content that aligns with a user's existing preferences, reinforcing existing belief clusters. Eventually, this contributes to a political landscape that is organized along a dominant axis, such as left versus right, collapsing what might otherwise be cross-cutting sources of identity and belief into a singular, polarized divide.

While it used to be possible for e.g. conservatives to care intensely about economic freedom but have moderate views on social issues, or conversely to have strict views on social issues but care about environmental conservation, the clustering of beliefs along the partisan dimension suppresses such cross-cutting ideological positions. Under polarization, conservatives and liberals adopt more rigid bundles of beliefs that align with their group identity, collapsing diverse preferences into a single dimension (Mason, 2018). Initially non-political belief systems like climate change or vaccination obediently align on that one dimension.

Seen through the lens of moral trade, such "dimensional collapse" restricts the trade space, and leads to a retrenchment of the Pareto frontier. Why is this, exactly? Recall that in the example of the vegetarian and the open-source software advocate, each agent's indifference on the other's "pet issue" represents their productive potential: it is what they can supply and what the counterparty to the trade will "consume," in the form of a commitment to align more closely on the counterparty's valued dimension.

Indeed, the following appears true: a population of moral agents, each valuing a unique "pet issue" exclusively and indifferent to all others, could use moral trade to achieve the same level of aggregate social welfare as a population that has reached moral consensus on a single shared issue. In each case, Self would get full alignment from every Other on Self's preferred moral dimension, since each Other would be willing to exchange commitments on Self's dimension, which they would be indifferent over, in exchange for Self making equivalent commitments on Other's preferred moral dimension.[8] Aggregate social welfare in equivalent across both cases, because each agent achieves the maximum utility they could derive within their moral framework: in the first population, through decentralized but perfectly efficient trade, and in the second population, through coordinated collective action. The first population relies on a network of reciprocal agreements and decentralized exchanges, while the second achieves maximum welfare through inherent alignment without the need for negotiation.

In sum, the existence of beliefs that do not align neatly along a single dimension that generate opportunities for moral trade, by preserving independent ideological parameters on which to trade. Thus, while differences in beliefs are at the basis of moral trade (Ord, 2015), when these differences exist on a single dimension, the possibility of trade relies on the existence of some orthogonal issue dimension that both moral agents are not opposed on. With rising polarization, such orthogonal issues become increasingly difficult to identify.

Other contemporary trends have the opposite effect, and we have already seen one instance in which this happens. As Norway grows wealthier, its moral scope may expand. While Norwegians used to care primarily about labor standards for Norwegians, they now come to care about the labor standards of Brazilian workers, perhaps because they are connected to those workers through conventional trade. In that case, they may gain from an increase in moral alignment on the part of Brazilian workers, as expressed through higher labor standards. If so, then the trade space grows as a result of the expansion of one actor's moral universe. As that actor comes to have preferences over the revealed preferences of foreign actors, moral trade becomes possible—which may come at the expense of conventional trading opportunities. As I suggest above, there is thus a potential substitution effect between moral trade opportunities and conventional trade opportunities.

**Maximizing Moral Welfare** In the case of conventional trade, the claim can be made that given gains from trade, exogenously increasing preference heterogeneity, in a way that allows for greater trade opportunities, might result in a net welfare increase.[9] As I argue above, the same does not follow in the case of moral trade: although its possibility improves outcomes at any level of non-zero preference heterogeneity, it does not entail that one would ever want to increase the

[8] The necessary assumption is that there are no barriers to trade, or transaction costs, and that all issues are orthogonal, meaning there is no overlap in preferences or values between individuals. An additional caveat might be that focusing collective efforts on a single dimension may have greater effects on that dimension that distributed commitments on as many dimensions as there are individuals. Consensus may thus have positive scale effects.

[9] This claim requires particular assumptions about intra-personal welfare comparisons. See fn 4, supra.

diversity of moral preferences to capture these gains.

One might ask, therefore, how would a benevolent dictator go about maximizing moral welfare? A first step would be try and achieve as much convergence between moral beliefs as possible, in a way that *reduces* the opportunities for moral trade. The second step would identify dimensions that do not correlate with clashing moral values. This is what the empirical section below begins with: using the General Social Survey (GSS), I identify those views are most weakly correlated with the partisanship dimension. These are beliefs around conservation and preservation of nature, human and animal health and welfare. Third, the benevolent dictator would maximize moral trades across the existing system of beliefs.

What this three-step process highlights is that unlike commodity trade, moral trade is only ever about closing the gap between divergent beliefs. In the best-case scenario, moral trade indeed transforms zero-sum conflicts into positive-sum solutions. Yet it does not follow that social welfare can increase with greater diversity in moral beliefs, as compared with moral consensus. To borrow terms frequently invoked in the discussion over climate change, we can ask what role moral trade can play in the mitigation of, and adaptation to, polarized societies. My main theoretical claim is that the moral trade framework can be thought of as an adaptive response to one of the central challenges of this political era. The reduction of polarization still represents a first-best outcome; yet once efforts to reduce polarization run their course, moral trade emerges as a singular mechanism to contend with the divisions that remain, by structuring clashing values into mutual moral utility gains. In the empirical section, I also examine the possibility that by fostering exchanges between moral adversaries, moral trade might also play a mitigating role, by reducing out-party animosity between opposed camps.

## 3 Empirics

In divided societies, the potential gains from moral trades are vast. Yet would such trades ever take place? This is the question I examine next. There are several obstacles to moral trade. Some are normative, and others arise from the structure of moral trade itself. I outline these obstacles, and examine the extent to which they affect individual attitudes towards moral trade across two survey studies. Given the article's motivation around polarization, the trade scenario I test empirically is one between opponents on a single dimension. I consider two different issues: political partisanship (Democrats v Republicans), and reproductive rights (pro-choice v pro-life).

### 3.1 Testable Expectations

**Preference Intensity** I am interested in how the strength of people's ideological stances on the underlying issue affects their attitudes to moral trades. My first pre-registered expectation is that all else equal, respondents with stronger stances should be less open to moral trades. Given what they know about themselves, stronger ideologues have more reason to believe that insofar as the counterparty is picked at random from the population, that counterparty will have relatively weaker ideological convictions. The respondent with more intense preferences would thus be committing to relatively more by abstaining from a donation to the preferred cause than their counterparty. Put otherwise, stronger ideologues have a steeper marginal rate of substitution between supporting their preferred cause and refraining from donating, meaning they require a disproportionately larger concession from the counterparty to justify their own sacrifice—yet here that concession is fixed by design. The trade would thus appear less attractive to more committed adherents to either issue.

We can capture this intuition, and derive other testable implications, by modelling the trade in the following way. Consider a moral trade scenario between a Democrat (Dem) and a Republican (Rep), each of whom wants to donate \$100 to their preferred political cause. They are offered a deal under which they could repledge their donation to a non-political public good—public park conservation—in exchange of the other doing the same. To see payoffs from the different cooperative outcomes, let:

$D_{\text{Dem}} \in [0, 100]$ denote the Democrat's donation to their own party, with $100 - D_{\text{Dem}}$ allocated to public park conservation.

$D_{\text{Rep}} \in [0, 100]$ denote the Republican's donation to their own party, with $100 - D_{\text{Rep}}$ allocated to public park conservation.

$\gamma_i > 0$ capture each player's ideological intensity.

$\alpha \in (0, 1)$ capture the marginal utility both players derive from the shared public good.

Each player's utility depends on three factors: (1) the benefit from donating to their own party, (2) disutility from the opponent's partisan donation, and (3) utility from the shared benefit of public park conservation. The utility functions for each player are thus:

$$U_{\text{Dem}}(D_{\text{Dem}}, D_{\text{Rep}}) = \gamma_{\text{Dem}} D_{\text{Dem}} - \gamma_{\text{Dem}} D_{\text{Rep}} + \alpha \left(200 - (D_{\text{Dem}} + D_{\text{Rep}})\right),$$

$$U_{\text{Rep}}(D_{\text{Dem}}, D_{\text{Rep}}) = \gamma_{\text{Rep}} D_{\text{Rep}} - \gamma_{\text{Rep}} D_{\text{Dem}} + \alpha \left(200 - (D_{\text{Dem}} + D_{\text{Rep}})\right),$$

Two expectations follow:

- As a player's ideological intensity increases ($\gamma_i \uparrow$), partisan utility dominates, making the moral trade less attractive.
- A more committed opponent, by symmetry, increases the incentive to cooperate in the moral trade.

If the Republican's ideological intensity is high ($\gamma_{\text{Rep}} \gg \alpha$), convincing them to abstain from donating imposes a higher internal cost on them, which amplifies the Democrat's relative gain. It follows that each party would have a potential incentive to misrepresent the intensity of their beliefs, to convince the counterparty they are getting a better moral deal, giving rise to concerns over credibility. The socially optimal outcome occurs when both cooperate. However, this outcome is unstable because unilateral defection offers a partisan advantage, particularly for players with high $\gamma_i$.

To test these expectations, both survey experiments elicit the respondent's partisan commitment. Study 1 also elicits each respondent's moral position on, and level of commitment to, reproductive rights. In Study 2, I then randomly vary the strength of the out-partisan, from "moderate" to "strongly committed".

**Affective Considerations** Through laboratory settings and online field experiments, the literature on affective polarization has shown that people appear less willing to work for, go into business with, purchase goods from, or date members of the opposing political party (Iyengar and Westwood, 2015). In this sense, affective polarization acts as a tax on social interaction (Kim and Pelc, 2024*a*). We might thus expect that respondents would be less likely to enter into any kind of trade—but perhaps especially a moral trade, which primes the partisan dimension—as the ideological distance between them and the trade partner increases. If so, then this would have implications for our expectations on preference intensity, above. Specifically, the more intense each party's ideological preferences, the greater the distance between the two, and the less likely either party would be willing to trade.

In other words, our tests on preference intensity amount to a type of horse-race between two explanations. In particular, if the counterparty's preference intensity increases respondents' openness to trade, then this would support formal expectations about moral gains. If it decreases openness to trade, then this would support an affective explanation, whereby in-partisans are keen to avoid interactions of any type with out-partisans.

**Deontologists vs Consequentialists** Moral philosophers and psychologists draw a broad distinction between two types of moral reasoning: deontological and consequentialist. Consequentialist reasoning focuses on the desirability of outcomes, while deontological reasoning (which approximates the Kantian approach) emphasizes the inherent rightness of the means employed to achieve those outcomes. Kantians are primarily concerned with moral motivations, and they view a gesture as morally right if it is performed out of moral duty. The premise is that these distinct moral models meaningfully track individual reasoning and behavior.

If moral behavior demands that gestures be taken for the right moral reasons, then Kantians might perceive moral trade as a sleight of hand. In this way, a pro-choice advocate does not merely want others to support reproductive rights as a matter of mutual arrangement; they believe all individuals should endorse this stance because it reflects principles of autonomy and equality that are normatively defensible. In other words, a deontological approach may be in tension with moral trade, where the reasons for others' ethical commitments are not based on the shared principle, but on a principle of exchange that might be driven by the combination of opposite principles (as with two moral agents refraining from donating to opposite causes from an understanding that these donations cancel each other out) or orthogonal principles (as with those two agents re-channeling their donations to a mutually tolerable third cause). More generally, deontologists typically view moral principles as holding universally, regardless of specific circumstances or reciprocal agreements. The conditional aspect of moral trade—where Self does something in exchange of Other doing something else—might be seen as clashing with this universalist approach. On the other hand, a Kantian might argue that moral trade does not undermine moral duties—it merely helps Other fulfill their moral duties more effectively, from Self's standpoint. Since moral trade involves a freely chosen exchange, a Kantian would recognize its respect for the other person's reason and agency. Finally, while the conditional structure of moral trade may seem to violate universality, the trade's outcomes promote universal moral ends, which may also be seen as aligned with Kantian values. As for consequentialists, they would see no inherent harm in the conditional aspect of trade, and would only judge of their desirability by how the end outcome affects overall welfare.

It is possible to test these competing expectations. To unpack the type of moral reasoning that people engage in when evaluating moral trades, I rely on a standard battery of questions commonly used by moral psychologists (Mata, Vaz and Mendonça, 2022) to rank individuals on a spectrum from consequentialist to deontological moral reasoning (see Appendix). Much of this literature builds on the trolley problem, a thought experiment in which sacrificing one life can save a number of other lives (Bruers and Braeckman, 2014). The resulting pre-registered expectation is that while both moral types may be open to moral trade, those respondents who are closer to a deontological moral perspective should be more resistant to it, while those who espouse a consequentialist view should be more open.

## 3.2 Survey Design

How open are people to moral trades, and what best explains their attitudes? To find out, I ran two studies. The first study recruited a quota-valid sample of 2,244 Americans through the survey firm Prolific. Respondents were selected to fill quotas on political partisanship, sex, and age. The survey was fielded from May 28th 2024 to April 4th 2024. The analysis was pre-registered on AsPredicted.[10] Given high rates of support for moral trade in the first study, a second study reran a second pre-registered experiment on partisan identity on a sample of 2,139 individuals, fielded from Nov 1st to Nov 4th 2024. This second study amended several aspects of the trade to make it less appealing, starting with its timing, in the final days of the 2024 Presidential campaign, when partisan emotions would be at their highest.

## 3.3 Moral Trade Issues: Partisanship and Reproductive Rights

**Trading Partisan Donations** The 2020 and 2024 US Presidential election saw record-breaking donations raised from individual donations. The most recent election saw a total of US$ 2.75 billion raised by both parties.[11] Most of this money was spent on television and digital ads. In past elections, a large proportion, ranging from 61% in 2012 to 34% in 2020, consisted of "attack ads."[12] Evidence from the study of electoral politics suggests that donations are often a response to donations to the other side. Well aware of this, political parties often exploit this dynamic to motivate donations to their camp. In 2012, a UCLA tax law professor attempted a real-life moral trade platform during the US presidential elections, called RePledge. He went as far as obtaining an opinion from the Federal Elections Commission, which eventually green lit the platform. Yet the experiment was never run, out of a lack of funds.[13] The purpose of this study is to examine who the likely participants of such an platform might have been, and how its design could have

[10] PAP available at `https://aspredicted.org/5CW_YQC`.

[11] Combining committee donations and outside donations. Statista, from Open Secrets data.

[12] https://mediaproject.wesleyan.edu/2020-summary-032321/

[13] According to the RePledge founder, the estimated cost of running such a platform in 2012 was USD$2 million, mostly to set up a credit card payment system. Current alternatives would arguably achieve this at a fraction of the cost.

encouraged exchanges between partisan foes.

### 3.4 Study 1

Study 1 was interested in how open people are to moral trades, what factors this depends on, and how malleable these attitudes are. There are two main outcome variables: respondents were asked (i) whether they would agree to a deal, and (ii) whether they thought such a deal would make the world better. The survey experiment had two separate treatments, resulting in four treatment arms. First, respondents were exposed to a randomized informational treatment, which aimed to offer an analogous instance of moral trades in the political arena. This consisted of a brief explanation of logrolling, where legislators trade votes to help the passing of a bill. The control group was exposed to an equally brief explanation of three steps involved in the passing of a legislative bill (see the Appendix for treatment formulation).[14]

Secondly, respondents were randomly assigned moral trade scenarios relating to financial donations in one of two issue-areas: (i) political partisanship (Democrat vs. Republican), and (ii) reproductive rights (pro-choice vs. pro-life). In each case, we first elicited the respondent's own stance on the issue, and then assigned the opposite stance to the moral trade partner, Taylor.[15] We also elicited respondents' preferred charity, from an option of three charities—in descending order of popularity: a children's hospital, an animal shelter, and an environmental conservation organization. These options were selected by looking the issues that correlated least highly with partisanship on the GSS survey: public health, animal welfare, and conservation.

All proposed moral trades thus consisted of pairings of one Democrat and one Republican, or one pro-choice and one pro-life respondent, who were offered the same deal: if you commit to abstaining from a donation to your ideologically preferred cause, the counterparty will do the same, and you will both channel the sum to a non-ideological charity. We elicited the intensity of the respondent's stance on their assigned issue (political partisanship or reproductive rights), to test *H1*.

[14] Respondents were asked a follow-up question to test their attention to treatment. In the case of the treatment: What is vote trading also known as? Logrolling (1), Filibustering (2), Gerrymandering (3).
In the case of the control: Which of the following is a crucial step in the legislative process before a bill can become law? Hearings (1), Filibustering (2), Gerrymandering (3).

[15] We chose the name Taylor for gender-neutrality, selecting from a list of the top ten names in the US with the most equal numbers of men and women.

## 3.5 Findings: Study 1

### Outcome Variable 1: Openness to Moral Trade

### Partisanship Trade

Imagine that you have a sum of $100 to allocate to a [Democratic/Republican] political cause of your choice. You meet Taylor, who intends to give $100 to a [Republican/Democrat] political cause. A common friend points out that you are both giving to opposite political causes. This friend suggests that you could both agree not to donate to your respective political causes, and each give the $100 to a cause you both support: a [children's hospital]. Under this arrangement, the [Democratic/Republican] organization would not get your $100. The [Republican/Democrat] organization would not get Taylor's $100. And the [Selected Charity] would receive both your donations, for a total of $200.

Would you be more likely to agree to, or reject this deal?

○ Agree to the deal: I would not donate my $100 to the [Democratic/Republican] cause, and I would agree to donate it to [selected charity] instead, on condition that Taylor does the same.
○ Reject the deal: I would prefer to donate my $100 to the [Democratic/Republican] cause, and let Taylor donate to the [Democratic/Republican] organization.
○ I don't know, I'm unsure what I would do.

### Reproductive Rights Trade

Imagine that you have a sum of $100 to allocate to a [pro-choice/pro-life] organization of your choice. You meet Taylor, who intends to give $100 to a [pro-choice/pro-life] organization. A common friend points out that you are both giving to opposite political causes. This friend suggests that you could both agree not to donate to your respective political causes, and each give the $100 to a cause you both support: a [selected charity]. Under this arrangement, the [pro-choice/pro-life] organization would not get your $100. The [pro-choice/pro-life] organization would not get Taylor's $100. And the [selected Charity] would receive both your donations, for a total of $200.

Would you be more likely to agree to, or reject this deal?

○ Agree to the deal: I would not donate my $100 to the [pro-choice/pro-life] cause, and I would agree to donate it to [selected charity] instead, on condition that Taylor does the same.
○ Reject the deal: I would prefer to donate my $100 to the [pro-choice/pro-life] cause, and let Taylor donate to the [pro-choice/pro-life] organization.
○ I don't know, I'm unsure what I would do.

Finally, in a follow-up question, I elicited respondents' reservations about agreeing to these deals. The survey proxied for resistance against the commodification of principled beliefs, beliefs about differential effectiveness, fears over a betrayal of personal principles, and doubts over the counterparty's credibility.

# 4 Study 1: Findings

Figure 1 shows the distribution of responses to the first outcome variable. The clear take-away is that a significant proportion of US respondents appear open to moral trades: 86.2% in the case of partisanship deals; 73% in the case of reproductive rights. This difference between the two issue-areas is statistically significant.

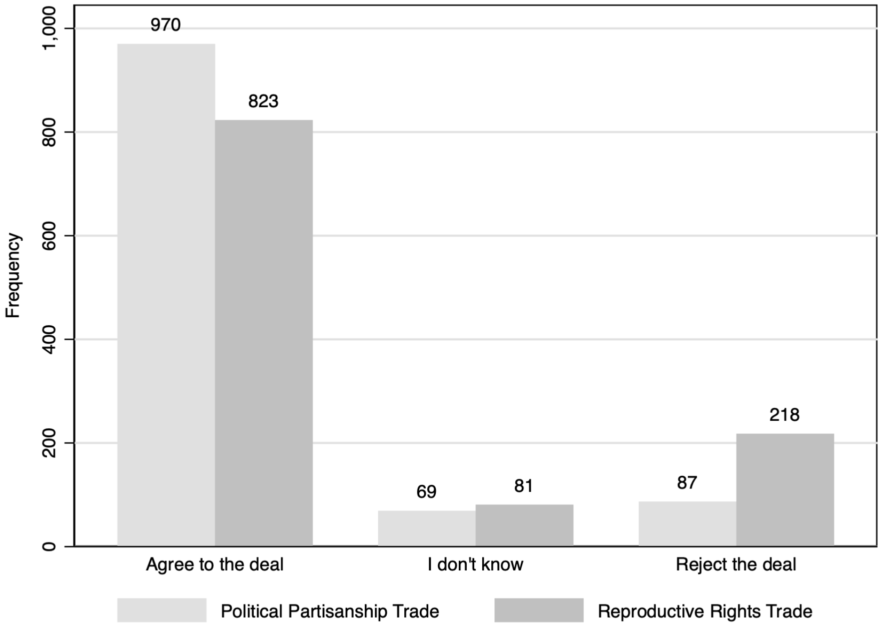

Figure 1: Study 1: Openness to Moral Trade Across Two Issue-Areas

**Determinants of Individual Attitudes** What drives variation in attitudes towards moral trades? We estimate a linear model of approval on the same outcome question as above. Our main variables of interest are the two treatments (informational treatment and issue-area).

To obtain a single score per respondent, we can then subtract the deontological score from the consequentialist score (these, as expected, as negatively correlated with each other, $\rho$= -0.35) The distribution of the net consequentialist score is shown in Figure 2. Based on this common battery of moral reasoning questions, most of our respondents fall on the deontological side of the spectrum. Few appear to be strict consequentialists.

The estimates control for several demographic traits, including sex, education, age, and income. Additionally, we test whether religious non-believers might hold different attitudes towards moral trades. The expectation is that religious believers may be more guarded about striking deals

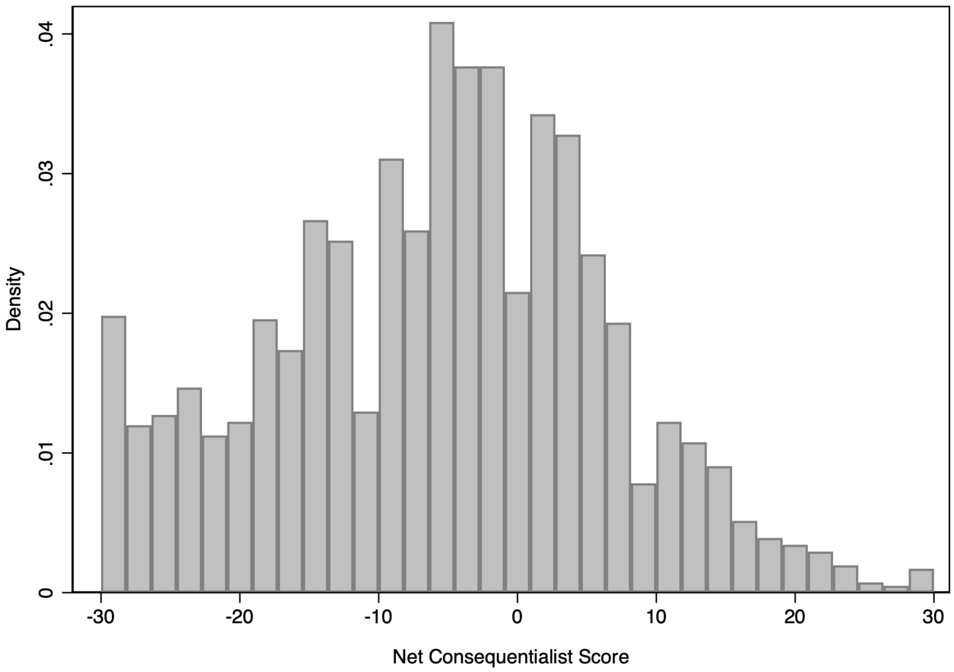

Figure 2: Distribution of Moral Reasoning Types

over normative positions, compared to non-believers.

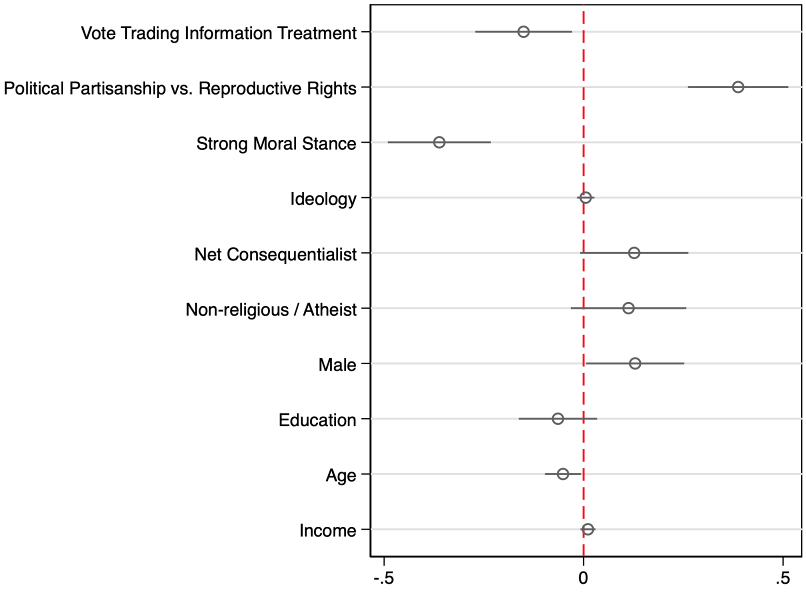

Figure 3: Determinants of Attitudes Towards Moral Trade
Estimates from a probit estimation of response item "Agree to the deal", n=2244.

Figure 3 shows the results. Several findings emerge. The first confirms the observation from Figure 1, which is that respondents are significantly more open to moral trades on political partisanship than on reproductive rights. The main surprise relates to the informational treatment. The initial preregistered expectation was that a mention of a practice analogous to moral trade in the US legislature would normalize moral trades, and thus *increase* respondents' openness. The actual result was exactly the opposite, at a statistically significant level: respondents exposed to information about logrolling in the US Congress became considerably more resistant to moral trades. This negative treatment effect proves stronger than either the gender or religious believer effect. One plausible interpretation is that rather than normalizing the notion of moral trades, the treatment may have foregrounded the latent reservations people have about moral compromise—triggering taboo tradeoff aversions and threatening sacred values, which could explain the backfiring effect. A large proportion of the sample recoiled from a move that, stripped of context, they might have found quite appealing on consequentialist grounds. If so, this seems consistent with some of the findings in Tetlock et al. (2000), where treatments that compromise so-called "sacred" values lead to greater moral cleansing, whereby lab participants doubled down on these values, in an attempt "to shore up the normative order" (860). Supporting this possibility, the negative treatment effect is heightened for net deontologists ($\beta$= -0.214, SE = 0.073, 95% CI [0-.358, -0.0701]), and disappears entirely for net consequentialists—those to the right of the zero-line in Figure 2.[16] In other words, those predisposed to judge actions based on duty or principle were especially turned off by the comparison to logrolling, while pure consequentialists were less perturbed.

Secondly, the expectation that individuals with stronger ideological stances on the underlying issue would be more resistant to moral trades is supported. This is relevant insofar as it is precisely such individuals who are most likely to be donors to their respective cause, and thus candidates for moral trades. Yet even among the strong moral stance group, moral trade garners majority support, with the average response at 6.48 ($\sigma$ = 2.64) on a 0 to 10 scale. This compares to a mean support of 7.16 ($\sigma$ = 2.45) among weaker moral stances on the issue.

Net consequentialists—that is, those who scored above 0 on an index that subtracted individual consequentialist scores from individual deontological scores—were more likely to support moral trades, but only at a 10% level of significance. By contrast, left-right political ideology was not associated with attitudes one way or the other. And non-believers/atheists appear more likely to support moral trade, this falls just short of statistical significance. Since atheism and ideology are strongly correlated ($\rho$=-0.35), we also include them separately, which does not qualitatively

[16] Treatment effect for the net consequentialists sample subset: $\beta$=.006 SE=0.116, 95% CI [-0.221, 0.232].

affect the coefficient size.

In terms of demographics, finally, younger male respondents appear more open to moral deals. On the other hand, education is, if anything, negatively associated, though it falls short of significance. Income sees no association in either direction, though it remains important as a control variable, since the questions concern financial donations, where income may play a role.

### Outcome Variable 2: Welfare Effects

> How much do you agree with the following statement:
> The world would be better off if you and Taylor agreed to the deal. [0-10 slider]
> Strongly disagree (0) — Strongly agree (10)

### Outcome Variable 3: Reasons Against Moral Trade

> Why might you think twice about agreeing to the arrangement your friend suggested, where you both agree not to give to the organization of your choice, and you both give to [Selected Charity] instead? Select your main concern.
>
> ○ I don't trust Taylor to keep their end of the deal. (1)
> ○ I think [Democratic/Republican, Pro-Life/Pro-Choice] organizations are more effective than [Republican/Democratic, Pro-Choice/Pro-Life] organizations. As a result, my donation will be more effective than the donation to the [Republican/Democratic, Pro-Choice/Pro-Life] cause. (2)
> ○ People should not make deals on the basis of political or moral beliefs. (3)
> ○ By agreeing to this deal, I would be betraying my beliefs. (4)
> ○ None of the above. (6)

**Factors of Resistance** We also elicited from respondents the reasons for which they might be resistant to moral trades. Table 1 shows the breakdown across the informational treatment, broken down by issue-area (political partisanship trades vs. reproductive rights trades). The main take-away is that overall, normative concerns appear to loom less large than practical ones. Fewer worry about either betraying their beliefs by making deals over moral beliefs, or normative concerns over making deals on the basis of moral beliefs. By contrast, there seems to be greater concern over the counterparty's credibility in keeping their end of the bargain, and some (lesser) belief in the greater effectiveness of donations to one's own cause than the counterparty's cause. Finally, the high proportion of respondents choosing "None of the above" may make best be thought of as a reflection of the high approval rates of moral trades across both issues areas. Accordingly, Partisanship Trade deals, which saw the highest approval also saw the highest proportion of "None of the above".

| | **Partisanship Trade** | **Reproductive Rights Trade** |
|---|---|---|
| 1. By agreeing to this deal, I would be betraying my beliefs. | 44 | 140 |
| 2. People should not make deals on the basis of political or moral beliefs. | 160 | 189 |
| 3. I don't trust Taylor to keep their end of the deal. | 305 | 256 |
| 4. I think [my] organizations are more effective than [opposed] organizations. | 136 | 155 |
| 5. None of the above. | 477 | 382 |
| **Total** | **1122** | **1122** |

Table 1: Factors Behind Resistance to Moral Trade

## 4.1 Study 2: Design

Given the surprisingly high level of support for moral trades in the first study, a follow-up study modified several aspects of the trade setting to make moral trades less appealing. First, the survey was run in the final days of the 2024 US Presidential campaign, when partisan emotions would be at their highest. Second, the sample was limited to self-described Democrats and Republicans, and omitted any independents, or partisans of other parties. The intent was to focus on true partisans who would have "moral value" to demand and to offer on the traded dimension of partisanship, rather than collecting a sample representative of the general population. Third, I randomly varied the intensity of the *counterparty's* beliefs, between a moderate vs. a strongly committed partisan of the opposite party. Finally, I picked an orthogonal belief that would be less likely to elicit strong convictions: conservation of public parks. Together, these amendments were intended to get more variation in attitudes, and further examine what structural aspects of moral trade were likely to generate resistance. The main question of interest was openness to trade, with response items "Agree", "Reject", or "I don't know, I'm unsure what I would do". The full question is reproduced below.

**[Agreement to Moral Trade]**

Imagine that you have a sum of $100 to allocate to the [your party]. You have been matched with someone else taking this survey named Alex. Alex describes themselves as a ”strongly committed [opposite partisan].” They want to donate their $100 to the [opposite party].

Since you would both donate to opposite causes, you and Alex are each being asked whether you would agree to the following arrangement:

Would you agree not to donate to the [your party], and instead give the $100 to a program for public park conservation, if Alex agreed to redirect their donation in the same way?

Under this arrangement, the [opposite party] would not get Alex's $100. The [your party] would not get your $100. And the public park conservation program would receive both your donations, for a total of $200.

○ Agree to the deal: I would not donate the $100 to the [your party], and I would donate it to the public park conservation program instead, on condition that Alex does the same.

○ Reject the deal: I would prefer to donate the $100 to the [your party], and let Alex donate to the [opposite party].

○ I don't know, I'm unsure what I would do.

The second question of interest in Study 2 was whether a successful moral trade might, by itself, change respondents' affect towards out-partisans. The "contact hypothesis" holds that intergroup interactions should breed more tolerance (Tajfel et al., 1979). Studies of contact with out-groups have mostly looked to ethnic diversity (Nathan and Sands, 2023), yet given growing attention to affective polarization, it has increasingly been applied to the partisan context. Some work has found support for an equivalent of the contact hypothesis, by generating positive interactions across political lines in laboratory settings (Levendusky and Stecula, 2021). These effects seem to depend on the nature of the interaction: positive interactions reduce out-group hostility, while negative out-party contact exacerbates it (Wojcieszak and Warner, 2020). In our case, we are interested in whether the successful "completion" of a hypothetical trade would affect people's sentiment towards the other side. We get at these in two ways: by probing political tolerance, and by eliciting interest in conducting an online task with the counterparty to the trade.

**[Political Tolerance]**

How much do you agree with the following statement:
I disagree with most of the [opposite party's] platform, but I think some of their points are valid.
○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

**[Online Task]**

Once you have completed this survey, how interested would you be in participating in an online task, working together with someone like [the counterparty]? This task would earn you $15.
○ Very interested
○ Somewhat interested
○ Neither interested nor disinterested
○ Somewhat disinterested
○ Very disinterested

All respondents received both questions, but the order in which they received them was randomized. Half of respondents saw these questions prior to being queried about the moral trade with the out-partisan, and the other half saw them after being queried about the moral trade, and told that the other side had "agreed" to the trade. The treatment effect consists in the difference between both groups in political tolerance, and willingness to take part in the only task with the counterparty.

### 4.2 Study 2: Findings

The first descriptive observation from Study 2 is that once more, the rate of openness to moral trade appears high. On average, across all treatment groups, just over 83% of respondents claimed they would accept the moral trade. How did these attitudes vary? The first parameter of interest was the ideological intensity of the respondent and the counterparty. Table 2 shows the rate of openness to moral trade across the various ideological conditions. One interpretation of these results is that the two competing pre-registered expectations cancel each other out. That is, the pursuit of individual moral gain pushes for higher rates of agreement to moral trade, while affective aversion to out-partisans pushes in the opposite direction.

|  | Opponent |  | Total |
|---|---|---|---|
| Respondent | Ideologically Moderate | Ideologically Strong |  |
| Ideologically Moderate | 0.89 (SD=0.32)<br>(n=490) | 0.87 (SD=0.34)<br>(n=523) | 0.88 (SD=0.33)<br>(n=1013) |
| Ideologically Strong | 0.77 (0.42)<br>(n=581) | 0.79 (0.41)<br>(n=545) | 0.78 (0.42)<br>(n=1126) |
| Total | 0.82 (0.38)<br>(n=1071) | 0.83 (0.38)<br>(n=1068) | 0.83 (0.38)<br>(n=2139) |

Table 2: Percentage of Respondents in Favor of Moral Trade Across Ideological Conditions: Average proportion, SD, and Frequencies

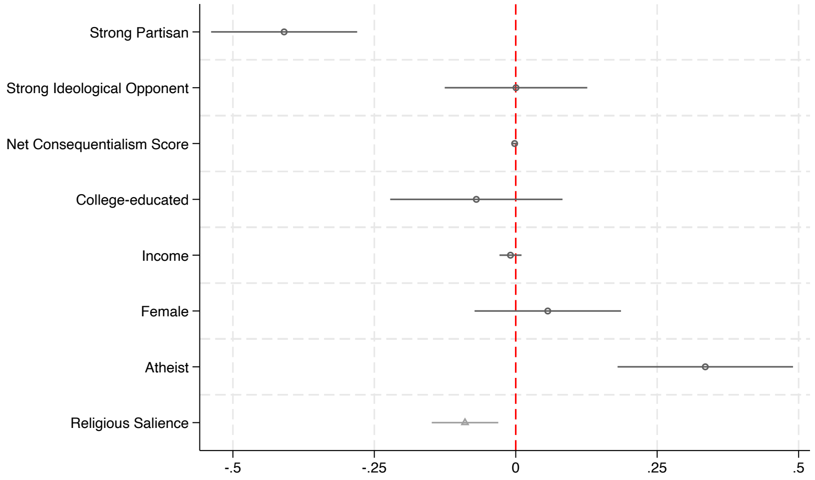

Figure 4: Determinants of Attitudes Towards Moral Trade
Estimates from an ordinal probit regression, response items “Reject the Deal”, “Don’t Know”, “Agree to the deal”, max sample n= 2139.

The second major question that Study 2 took up was whether a successful trade would change respondents’ aversion to the other side. Here, the results suggest a positive association, but one that falls short of statistical significance. Looking at heterogeneous effects, moderate partisans appear to be more swayed. It may be that a hypothetical moral trade with a hypothetical out-partisan is too weak of a treatment to observe an effect. Yet on the basis of these findings, one cannot conclude that the consummation of moral trades, by itself, brings ideological foes closer together. A fuller examination would ideally involve a real world iterated setting, where out-partisans could observe for themselves the result of their commitment in the counterparty’s behavior, and adjust subsequent interactions with counterparties accordingly.

Table 3: Agreement to Moral Trade on Political Donations

| | (1) | (2) | (3) |
|---|---|---|---|
| Strong Partisan | -0.410*** | -0.409*** | -0.470*** |
| | (0.065) | (0.066) | (0.076) |
| Strong Ideological Opponent | 0.008 | 0.000 | 0.035 |
| | (0.064) | (0.064) | (0.073) |
| Net Consequentialism Score | | -0.002 | -0.002 |
| | | (0.003) | (0.003) |
| College-educated | | -0.070 | -0.071 |
| | | (0.078) | (0.089) |
| Income | | -0.009 | -0.004 |
| | | (0.010) | (0.011) |
| Female | | 0.057 | 0.076 |
| | | (0.066) | (0.075) |
| Atheist | | 0.335*** | |
| | | (0.079) | |
| Religious Salience | | | -0.090*** |
| | | | (0.030) |
| cut1 | -1.296*** | -1.296*** | -1.581*** |
| | (0.062) | (0.111) | (0.157) |
| cut2 | -1.168*** | -1.167*** | -1.460*** |
| | (0.060) | (0.111) | (0.156) |
| Observations | 2139 | 2139 | 1582 |

Dependent variable is likelihood of agreement to moral trade with an out-partisan. Estimates from an ordinal probit regression, response items "Reject the Deal", "Don't Know", "Agree to the deal". * $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Model 3 excludes self-described atheists.

# 5 Conclusion

This article is a first attempt at formalizing the properties of moral trade, and testing some empirical expectations around its viability. I define a concept of moral utility, and argue that the distinctive quality of the moral setting is the interdependent nature of individual preferences: each actor derives utility from the revealed preferences of every other actor, on their moral dimensions of interest. I then discuss what the fundamental concepts of marginal rates of substitution, cost ratios, and comparative advantage represent in the moral setting, and how they differ from the conventional trade setting. Moral trade allows us to apply trade theory to an area where it has not previously been applied, with fruitful results. I suggest that moral trade may in fact describe social phenomena that are already widespread: moral trades may thus be the best means of conceiving of modern trade agreements, where some countries trade market access in exchange for commitments on values ranging from human rights and environmental sustainability to labor standards.

In the second half of the paper, I turn to individual attitudes towards moral trade, by asking whether such trades would ever take place. The short answer is yes: supermajorities appear open to reaching moral deals with ideological foes on such fraught issues as reproductive rights, and at such sensitive times as the last days of a contentious Presidential campaign. On the basis of the survey results, moral trade holds promise as a welfare-enhancing mechanism in polarized societies.

To borrow the distinction between mitigation and adaptation from the discussion over climate change, the moral trade framework can be thought of as an adaptive response to one of the central challenges of this political era. As I demonstrate, the direct reduction of polarization remains a first-best outcome, even when moral trade is frictionless; yet once efforts to reduce polarization have run their course, moral trade emerges as a singular mechanism to contend with the divisions that remain, by structuring clashing values into mutual gains in “moral utility”.

Ultimately, the concept of moral trade matters less as a template for a future global moral trade summit—though the possibility beckons—than as a stylized version of a basic form of social interaction in the face of clashing beliefs. What formalizing the process lets us see is that these divisions are inherently costly, but they can be bridged in part through structured arrangements. It is what we might hope of well-disposed individuals who “agree to disagree”: at best, these clashes can be restructured in a way that is mutually beneficial. As I show, polarization and social sorting limits the moral trade space, and the likelihood of such mutually beneficial arrangements. Such exchanges are driven by the existence of cross-cutting identities; as individuals increasingly cluster around a single ideological dimension, the potential for moral trades disappears.

# References

Bettarelli, Luca, Andres Reiljan and Emilie Van Haute. 2023. "A regional perspective to the study of affective polarization." *European Journal of Political Research* 62(2):645–659.

Birdsall, Nancy, William Savedoff and Frances Seymour. 2014. "The Brazil-Norway agreement with performance-based payments for forest conservation: successes, challenges, and lessons." *CGD Climate and Forest Paper Series* 4.

Brady, Henry E and Thomas B Kent. 2022. "Fifty years of declining confidence & increasing polarization in trust in American institutions." *Daedalus* 151(4):43–66.

Bruers, Stijn and Johan Braeckman. 2014. "A review and systematization of the trolley problem." *Philosophia* 42:251–269.

Charnovitz, Steve. 1991. "Exploring the environmental exceptions in GATT Article XX." *J. World Trade* 25:37.

Charnovitz, Steve. 2005. "An analysis of Pascal Lamy's proposal on collective preferences." *Journal of International Economic Law* 8(2):449–472.

Duskin, Kayla, Joseph S Schafer, Jevin D West and Emma S Spiro. 2024. Echo Chambers in the Age of Algorithms: An Audit of Twitter's Friend Recommender System. In *Proceedings of the 16th ACM Web Science Conference*. pp. 11–21.

Flaxman, Seth, Sharad Goel and Justin M Rao. 2016. "Filter bubbles, echo chambers, and online news consumption." *Public opinion quarterly* 80(S1):298–320.

Green, Donald, Bradley Palmquist and Eric Schickler. 2008. Partisan hearts and minds. In *Partisan Hearts and Minds*. Yale University Press.

Huber, Gregory A and Neil Malhotra. 2017. "Political homophily in social relationships: Evidence from online dating behavior." *The Journal of Politics* 79(1):269–283.

Inglehart, Ronald F. 2013. Changing values among western publics from 1970 to 2006. In *European Politics*. Routledge pp. 130–146.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideology: A social identity perspective on polarization." *Public opinion quarterly* 76(3):405–431.

Iyengar, Shanto and Sean J Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science* 59(3):690–707.

Kim, Sung Eun and Krzysztof Pelc. 2024*a*. "Does Political Diversity Inhibit Blood Donations?" *Perspectives on Politics* pp. 1–21.

Kim, Sung Eun and Krzysztof Pelc. 2024*b*. "Taking One for the (Other) Team: Does Political Diversity Lower Vaccination Uptake?" *Political Behavior* pp. 1–21.

Levendusky, Matthew S and Dominik A Stecula. 2021. *We need to talk: how cross-party dialogue reduces affective polarization*. Cambridge University Press.

Liu, Xi, Clio Andris and Bruce A Desmarais. 2019. "Migration and political polarization in the US: An analysis of the county-level migration network." *PloS one* 14(11):e0225405.

Mason, Lilliana. 2018. "Ideologues without issues: The polarizing consequences of ideological identities." *Public Opinion Quarterly* 82(S1):866–887.

Mata, André, André Vaz and Bernardo Mendonça. 2022. "Deliberate ignorance in moral dilemmas: Protecting judgment from conflicting information." *Journal of Economic Psychology* 90:102523.

Millgram, Elijah. 1997. *Practical induction.* Harvard University Press.

Nathan, Noah L and Melissa L Sands. 2023. "Context and Contact: Unifying the Study of Environmental Effects on Politics." *Annual Review of Political Science* 26.

Ord, Toby. 2015. "Moral trade." *Ethics* 126(1):118–138.

Oxtoby, David W, Henry E Brady, Tracey L Meares, Lee Rainie and Kay Lehman Schlozman. 2023. "Distrust, Political Polarization, and America's Challenged Institutions." *Bulletin of the American Academy of Arts and Sciences* 76(3):42–64.

Paolini, Stefania, Jake Harwood and Mark Rubin. 2010. "Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures." *Personality and Social Psychology Bulletin* 36(12):1723–1738.

Reiljan, Andres. 2020. "'Fear and loathing across party lines'(also) in Europe: Affective polarisation in European party systems." *European journal of political research* 59(2):376–396.

Rodrik, Dani. 2018. "What do trade agreements really do?" *Journal of economic perspectives* 32(2):73–90.

Samuelson, Paul A. 1948. "International trade and the equalisation of factor prices." *The Economic Journal* 58(230):163–184.

Tajfel, Henri, John C Turner, William G Austin and Stephen Worchel. 1979. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56(65):9780203505984–16.

Tetlock, Philip E, Orie V Kristel, S Beth Elson, Melanie C Green and Jennifer S Lerner. 2000. "The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals." *Journal of personality and social psychology* 78(5):853.

Wojcieszak, Magdalena and Benjamin R Warner. 2020. "Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact." *Political Communication* 37(6):789–811.

# 6 Appendix

### Randomized Information Treatment

Read this text carefully. You will be asked a question about it on the following page:

**Logrolling Treatment**:

Vote trading, also known as logrolling, is a long-standing practice in the US Congress, whereby members agree to vote for each other's proposed legislation. It is a way for legislators to help ensure that the bills they support have enough votes to pass.

**Control:**

The legislative process in the US Congress involves a series of steps before a bill can become law. These steps include the introduction of a bill, hearings, and amendments. Each stage is crucial to ensure that the democratic process is respected.

### Consequentialist v Deontological Moral Reasoning Types

We will now ask you a few questions about your values and beliefs. Please tell us how much you agree with the following statements.
0 means that you "strongly disagree", 10 means that you "strongly agree".

○ It is never justified to cause harm or suffering to anyone.
○ A person's life is sacred, and killing is always wrong.
○ The act of lying is a violation of trust and duty, and it cannot be justified in an attempt to avoid harming others.

○ If causing harm or suffering to a person makes it possible to achieve greater good for a greater number of people, then it is justifiable.
○ If sacrificing one person means saving many more people, then it is permitted.
○ The act of lying is permissible if it prevents harm to others.