

---

# Fair Share? Equality and Equity in American Attitudes Toward Trade

Ryan Brutger  and Brian Rathbun 

---

**Abstract** American politicians repeatedly and strenuously invoke concerns about fairness when pitching their trade policies to their constituents, unsurprisingly since fairness is one of the most fundamental and universal moral concepts. Yet studies to date on public opinion about trade have not been designed in such a way that they test whether fairness is important, nor whether the mass public applies fairness standards impartially. Drawing on findings in social psychology and behavioral economics, we develop and find evidence for an “asymmetric fairness” argument. In a national survey of Americans, we find strong evidence that fairness, conceived in terms of equality, is crucial for understanding support for potential trade deals and support for renegotiating existing ones. Americans view as most fair and most preferable outcomes in which concessions and benefits are equal across countries, especially when those equal benefits match productivity. However, we find that Americans have an egoistically biased sense of fairness, responding particularly negatively to any outcome that leaves the United States relatively worse off—a sense of injustice that does not extend to the same degree to relative gains for Americans.

---

Behavioral economists have long made reference to fairness as a crucial factor in explaining human decisions about economic matters. Fairness is one of the most universal of moral norms, and it is difficult to imagine that it does not inform the opinions and judgments of individuals in matters of foreign economic policy. When American presidents talk about free trade, they are likely to mention the importance of fairness. But why? If economics is the way by which individuals increase their material well-being and foreign economic policy is the way by which governments increase the prosperity of their citizenry, then the standard for a good or bad trade deal should be whether any trade policy benefits national welfare.

The answer is likely political, imposed by domestic audiences. Finlayson and Zacher write that “most observers argue that reciprocity is a requirement imposed on trade negotiators by fundamentally political imperatives. Governments feel compelled to justify their tariff ‘concessions’ to instinctively mercantilistic domestic audiences by pointing out that major trading partners have made at least equivalent ‘sacrifices.’”<sup>1</sup> We agree. However, it remains an open question why the mass public would care about reciprocity, provided that the gains were high enough. Fairness seems to provide the answer.

1. Finlayson and Zacher 1981, 575.

In this research, we draw on two of the most prominent behavioral models—inequity aversion and reciprocal fairness—to build an argument about the trade preferences of ordinary Americans. We suspect that the public applies the same moral judgments to economic foreign policy as they do to interpersonal relationships. Behavioral economic models suggest that rather than making choices that are best for them in absolute terms, individuals' preferences are contingent on their relative share and the behavior of their partners. Both reciprocal fairness and inequity aversion models predict a commitment to the same moral principle: a general preference for mutual and equal gain with a simultaneous insistence on not being exploited. Consistent with this, our analysis of national surveys of Americans finds that people are relatively indifferent to the absolute size of trade concessions made to the United States, but are quite responsive to the equality of those concessions and resulting trade balances.

However, Americans' judgments are not based on an impartial application of the principle of fairness. Instead, fairness judgements are asymmetric in a partially self-serving fashion. Building on psychological research, we offer and find support for an asymmetric fairness argument, which expects that Americans will judge outcomes in which they come out relatively behind as considerably less fair than outcomes in which they come out relatively ahead. Just because fairness matters does not mean it does so in a way that would make a moral philosopher proud. This fairness bias does have its limits though, in that Americans judge equal outcomes as considerably fairer than relatively advantageous ones.

In a second survey, we find a similar pattern of results for scenarios in which respondents are asked to form judgments about whether to renegotiate an existing trade agreement based on information about the trade balance. Americans are particularly inclined to revisit (and find unfair) deals that leave them relatively behind, but they view balanced outcomes as fairer than advantageous ones. This commitment to fairness as equality is robust to the inclusion of another conception of fairness in the research design—equity, the matching of inputs to outputs. Regardless of whether foreign workers are more efficient, Americans dislike, and find unfair, trade outcomes that leave the United States relatively behind. Nevertheless the most fairly judged outcome of all is a balanced trade flow in which workers are equally productive, with each side receiving a fair share for their work.

Our study adds to the existing literature in a number of important ways. First, while it is now well established that Americans make judgments about trade based on socio-tropic considerations,<sup>2</sup> we address a void in the literature by designing our studies to assess how fairness shapes attitudes toward trade. Second, while international relations scholars have long turned to reciprocity as an important driver of international cooperation,<sup>3</sup> they have been less attentive to reciprocity's status as a moral norm of fairness. Behavioral economists conceive of fairness as a "social preference" in which

2. Bechtel and Liesch 2020; Guisinger 2017; Mansfield and Mutz 2009.

3. Bagwell and Staiger 1999; Bechtel, Genovese, and Scheve 2019.

individuals care not only about the gains to themselves but also to others, and not always in an altruistic manner. Reciprocity dictates that we punish others for defecting because they did wrong and reward them for cooperating because they did right. In international relations, reciprocity is generally thought of as a solely (although mutually) self-serving norm in which decision makers care about the decisions and outcomes for others only insofar as they affect themselves. Although the repeated nature of interactions makes it hard to distinguish between these two theoretical possibilities, our studies provide initial evidence of the plausibility of the former as an important supplement to the latter in international relations. Third, while others have established that fairness matters in international relations,<sup>4</sup> they have not been attentive to systematic bias—that is, the asymmetry in fairness judgments and how far it extends.

### **Behavioral Economics, Social Preferences, and Fairness**

Human beings rarely consider any outcome with distributive consequences without reference to fairness. Concerns about fairness—and its inextricable cousin, reciprocity—are thought to be so common as to be universal norms<sup>5</sup> and likely have biological origins in our evolved psychology. Fairness has been a central concern of behavioral economists for several decades; they find it to be a powerful predictor of human behavior.<sup>6</sup>

There are two prominent models of fairness: inequity aversion and reciprocal fairness. Inequity aversion models show that individuals generally prefer equal divisions of resources to those that favor them or that leave them behind. “They are altruistic towards other persons, i.e., they want to increase the other persons’ material payoffs, if the other persons’ material payoffs are below an equitable benchmark, but they feel envy, i.e., they want to decrease the other persons’ payoffs, when the payoffs of the others exceed the equitable [i.e., equal] level.”<sup>7</sup> Individuals define and prefer fairness as equal outcomes for all. Lü, Scheve, and Slaughter find evidence for this phenomenon in a trade setting, however one that looks at preferences for the distribution of gains domestically.<sup>8</sup>

Reciprocal fairness arguments define fairness in terms of reciprocity. As Rabin, who developed the model, colorfully puts it: “If somebody is being nice to you, fairness dictates that you be nice to him. If somebody is being mean to you, fairness allows—and vindictiveness dictates—that you be mean to him.”<sup>9</sup> In this model, individuals condition their behavior on that of others. Fairness tells us how we respond to the actions of others.

4. Gottfried and Trager 2016; Kertzer and Rathbun 2015.

5. Berg, Dickhaut, and McCabe 1995.

6. Fehr and Fischbacher 2002; Rabin 1998.

7. Fehr and Fischbacher 2002, C3.

8. Lü, Scheve, and Slaughter 2012.

9. Rabin 1993, 1281.

Both models, despite their differences, are actually capturing commitment to the same moral principle: a general preference on the part of individuals for mutual and equal gain with a simultaneous insistence on not being exploited. In the reciprocal fairness model, “whether an action is perceived as kind or hostile depends on the fairness or unfairness of the consequences and the intention associated with the action. The fairness of the intention, in turn, is determined by the equitability of the payoff distribution, relative to the set of feasible payoff distributions, caused by the action.”<sup>10</sup> In other words, inequity aversion is built into the model. Inequality is the mean behavior that is punished through negative reciprocity. Similarly, the findings of ultimatum games, taken as empirical demonstrations of the inequity aversion model, show that when others treat us unfairly by offering a division of resources that is unequal, we refuse them. Again, negative reciprocity is how we punish inequality.

Both reciprocal fairness and inequity aversion are examples of what Fiske calls “matching”: individuals match the behavior of others (reciprocity) to arrive at a matching outcome (equality).<sup>11</sup> This makes reciprocity and fairness different from altruism, a “form of unconditional kindness.”<sup>12</sup> Most, however, are not so unreservedly giving. Instead, Rabin observes, “the same people who are altruistic to other altruistic people are also motivated to hurt those who hurt them.”<sup>13</sup>

International relations scholars have noted the importance of reciprocity for international cooperation for a long time.<sup>14</sup> However, their conception of reciprocity is different than what prevails in behavioral economics in a crucial respect: it is predicated entirely on self-interested behavior. It is to individuals’ advantage to reciprocally cooperate (or reciprocally defect) when they can expect to interact with others on a repeated basis. Axelrod and Hamilton begin their classic book, *The Evolution of Cooperation*, with the central question: “Under what conditions will cooperation emerge in a world of egoists without central authority?”<sup>15</sup> In this central text, as well as in Keohane’s classic treatment of reciprocity, neither the concept of fairness nor the word itself is mentioned at all.<sup>16</sup>

10. Fehr and Fischbacher 2002, C3.

11. Fiske 1991.

12. Fehr and Fischbacher 2002, C3.

13. Rabin 1993, 1281. However, reciprocal fairness and inequity aversion models do sometimes make different predictions in that inequity aversion is outcome based while reciprocal fairness is intention based. In inequity aversion, grounded as it is in loss aversion theories (Fehr and Schmidt 1999, 824), fairness is determined by the overall outcome (albeit in a comparative fashion in which losses are defined in relative terms, unlike in prospect theory). In reciprocal fairness, the intentions of others matter. For instance, in ultimatum games, receivers are less likely to reject inequitable outcomes if the offer is made by some random process, as opposed to the other participant, which they would not do if intentions did not matter. From a reciprocal fairness model, we do not punish others for unequal outcomes unless they are intentionally being unfair. Blount 1995. However, our research design focuses on the negotiation concessions and outcomes of trade agreements, as opposed to the intentions of states, so we cannot distinguish between these two different models of fairness.

14. Frank et al. 2018.

15. Axelrod and Hamilton 1981, 3.

16. Keohane 1986.

As understood in the behavioral economics literature, fairness is a “social preference” in which “the person not only cares about the material resources allocated to her but also cares about the material resources allocated to relevant reference agents.”<sup>17</sup> Social preferences can be mean or generous, but they are not indifferent as is the case in typical models of cooperation in international relations in which the behavior of others matters only insofar as it affects one’s own payoffs. Unreciprocated cooperation harms us in absolute terms. Fairness is an inherently relational concept, however. How we feel about our gains depends on what the other got or did.<sup>18</sup> Behavioral economists distinguish their models from more canonical micro-economic approaches that assume pure egoism.<sup>19</sup> This conception is only recently finding root in international relations scholarship.<sup>20</sup>

Empirically of course, it is very difficult to distinguish between reciprocity driven by a sense of fairness and pure self-interest, which is one reason why behavioral economists have taken to the lab. They rely extensively on one-shot interactions in which, for instance, substantial numbers of individuals cooperate in one-round prisoner dilemma games with the expectation of reciprocity by others, or punish others for unequal offers in ultimatum games. This is the case even though there is no chance of building a reputation that would benefit long-term self-interest in either scenario.

However, to date research on trade preferences has not been set up to gauge the effect of fairness at all. For instance, Mutz and Kim find that Americans equally support agreements in which (1) the US gains and the partner country loses and (2) the US gains and the trading partner also gains, so they conclude that Americans are primarily interested in egoistic gains for their country rather than relative gains.<sup>21</sup> However, the framework does not allow them to test whether fairness considerations, a different type of relational standard, might matter. For instance, are Americans more supportive of a deal that leaves them less well off absolutely but provides equal gains to both sides than they are of a deal that leaves them better off absolutely but relatively behind?

### **Bias and Partiality in Fairness Judgments**

Even if there is reason to believe that individuals have a preference for fairness, it is far from clear whether that sense of fairness is not, at least partially, self-serving. If fairness is entirely impartial, determined by the outcomes or actions of both parties

17. Fehr and Fischbacher 2002, C2.

18. This evokes the “gains debate” between neorealists and neoliberals about whether state interests are predominantly relative or absolute in nature, as discussed by Grieco 1988. However, as Powell 1991 demonstrated, relative gains concerns of this kind are ultimately egoistic in origin. Relative gains concerns are not the expression of a social preference but rather egoism since the gains of others give them the possibility to do harm to one’s future egoistic ends through the use of coercion.

19. Fehr and Fischbacher 2002; Rabin 1993, 1281.

20. Kertzer and Rathbun 2015.

21. Mutz and Kim 2017.

rather than one's own particular position, then individuals, even if they prefer an egalitarian outcome, will be as disappointed with one that favors themselves as one that favors others. What is good for the goose is good for the gander. This is the standard for fairness used in Herrmann, Tetlock, and Diascro's study on trade and the strongest effect that fairness might have.<sup>22</sup> However, this is not what empirical research suggests will be the case.

Inequity aversion models distinguish between two types of inequality. Individuals exhibit strong aversion to "disadvantageous inequality," that is, outcomes that leave them relatively behind. However, they are more supportive of equality than "advantageous inequality"—outcomes that leave them ahead. In experiments that manipulate the gains from allocative games, Loewenstein, Thompson, and Bazerman find highest support for the equal outcome, which is also judged the fairest.<sup>23</sup> However, peoples' distaste for outcomes that leave them relatively worse off is much stronger than their distaste for outcomes that leave them relatively better off. If individuals applied the rule of fairness impartially, this would not be the case.

Research in social psychology suggests such a self-serving process in determinations of fairness. When placed, for instance, in a hypothetical bargaining setting based on a real-life court case in which individuals are randomly assigned to plaintiff and defendant roles and given the same information, participants tend to see their own side as having a better case, predict the judge will rule in their favor, and regard a fair outcome as one that tilts in their favor.<sup>24</sup> Pinker calls this the "moralization gap."<sup>25</sup> Even though human beings are moral animals and ethical rules are by nature supposed to be applied impartially, in fact people apply them unwittingly in partially (although not entirely) self-serving ways.

We suspect that this same partiality in fairness judgment plays a role in explaining the difference in how individuals judge disadvantageous and advantageous inequality, something not established in prior research. We call this an "asymmetric" conception of fairness—one that is subjective and particularly sensitive to slights to oneself, but not so biased as to simply judge fairness after the fact based on whether one came out ahead. In this conception, individuals are self-involved but not purely self-interested. They are not blind to fairness, but near-sighted, better capable of seeing injustices done to themselves than to others. We suspect that fairness, given its centrality to human life and cooperation, will inform the attitudes of the general public when it comes to their country's interaction with other countries. In other words, they will prefer fairness to govern interstate relations, just as it does their interpersonal relations, and they will demonstrate the same asymmetry in its application.

22. Herrmann, Tetlock, and Diascro 2001.

23. Loewenstein, Thompson, and Bazerman 1989.

24. Babcock and Loewenstein 1997; Babcock et al. 1995.

25. Pinker 2012, 490.

## Trade Concessions Study

Based on previous work we take as our starting point the general American-centric, sociotropic quality of American trade preferences, in which the mass public is generally interested in how the United States fares.<sup>26</sup> A number of studies have now established that trade attitudes cannot be reduced to just personal pocketbook considerations. Rather, Americans seem to judge trade agreements by how they affect the United States as a whole and ingroup/outgroup dynamics play a considerable role.<sup>27</sup> Nevertheless, fairness has been left out of these studies. We start from the premise that in the population as a whole, Americans think about what is fair in view of what is fair to the United States. And they do so in the same way that they approach fairness in their interactions with others, that is, asymmetrically.

To evaluate the significance and conception of fairness with regard to international trade, we introduce two survey experiments meant to capture the importance that the American public places on fairness when thinking about trade. We use survey experiments to test our theory since they allow us to isolate the causal effect of key components of our theory and evaluate what aspects of international trade agreements alter perceptions of fairness and support. The surveys were fielded with Survey Sampling International (SSI) in the fall of 2017 on a sample of Americans that is broadly representative based on demographics such as age, education, income, and gender (see appendix for recruitment information and sample characteristics).<sup>28</sup> The studies were fielded as part of an omnibus panel study conducted in late 2017, with 3,136 respondents completing our first experiment and 3,201 completing our second experiment.

Our first study tests how changes in the concessions made by each party to a trade negotiation affect perceptions of fairness and support for the trade agreement. Unequal tariff rates are key complaints in the campaign against unfair trade, and thus we vary the tariff concessions made by the US and another country. We randomly varied the concessions each side made, such that each side could make a 30 percent, 60 percent, or 90 percent cut to their tariffs, resulting in nine treatment combinations.<sup>29</sup>

26. Mansfield and Mutz 2009.

27. Mutz and Kim 2017.

28. For a sample of published political science studies using SSI, please see Berinsky, Margolis, and Sances 2014; Brutger and Kertzer 2018; and Kertzer and Brutger 2016.

29. We recognize that using tariff concessions has two primary limitations. The first is that some respondents may struggle to interpret percentages, which would bias against finding a result, making this a relatively hard test for our theory. The second is that the scenario does not provide tariff baselines, which we find is generally consistent with how leaders discuss tariff cuts, but nonetheless may limit respondents' ability to evaluate the relative fairness of the concessions. We address both of these concerns in our second study, which has the advantage of presenting the relative relationship in terms of trade balances, as opposed to percentages of tariff cuts, which is perhaps more easily understood by respondents and avoids any complications of what the initial baseline of the tariffs might have been.

The text of the first experiment is as follows:

The US is considering negotiating a trade agreement with one of its trading partners. The trade agreement will decrease the average tariffs—that is, the tax charged by the American government on foreign goods entering the United States—by [30, 60, *or* 90] percent. In return the trade partner will decrease their tariffs on imports from the US by [30, 60, *or* 90] percent.

After reading about the trade agreement, participants were then asked whether they would support or oppose the agreement, with responses ranging from “strongly support” to “strongly oppose” on a five-point scale. Respondents were also asked to consider how fair they thought the trade agreement was, with responses ranging from “very unfair” to “very fair” on a five-point scale.

For our hypotheses, we group these nine treatments into three broad categories based on whether the tariff concessions are *equal*, *favorable*, or *unfavorable*. The equal category comprises those in which parties make the same concessions, whereas in the favorable treatment the other country makes a larger concession than the US, and in the unfavorable treatment the US makes a larger concession than the other country.<sup>30</sup> This means that the equal category of treatments includes equal concessions of 30/30, 60/60, and 90/90 by both sides. The favorable category of treatments includes concessions of 30/60, 60/90, and 30/90, where the first number is the percent tariff cut by the US and the second is the percent cut by the other country, and the unfavorable category of treatments is the inverse of the favorable treatment.<sup>31</sup>

A simple sociotropically egoistic account of trade preferences would expect that support for trade agreements would generally rise and fall based on the country’s gains.

*H1: Support for trade deals will be a function of absolute gains for the United States.*

In contrast, our asymmetric fairness argument has two primary expectations. First, Americans will judge trade outcomes with a relational standard, disliking disadvantageous inequality when it comes to trade outcomes but not favoring advantageous inequality to the same degree. In other words, they are more concerned with falling behind than coming out ahead. If they apply the same metric as they do interpersonally in other studies, they might even favor equal outcomes over favorable ones. Second, their conception of what is fair will be asymmetric. Americans will be more sensitive to the unfairness of outcomes that leave them behind than those that are disadvantageous to others.

30. We recognize that from an economic perspective where trade liberalization can create gains from trade, asymmetric reductions in tariffs may not necessarily be economically favorable or unfavorable, however, we use these terms since they are consistent with the political rhetoric surrounding the issue.

31. In the appendix, we conduct balance tests that demonstrate our random assignment achieved a well-balanced study across treatment conditions.

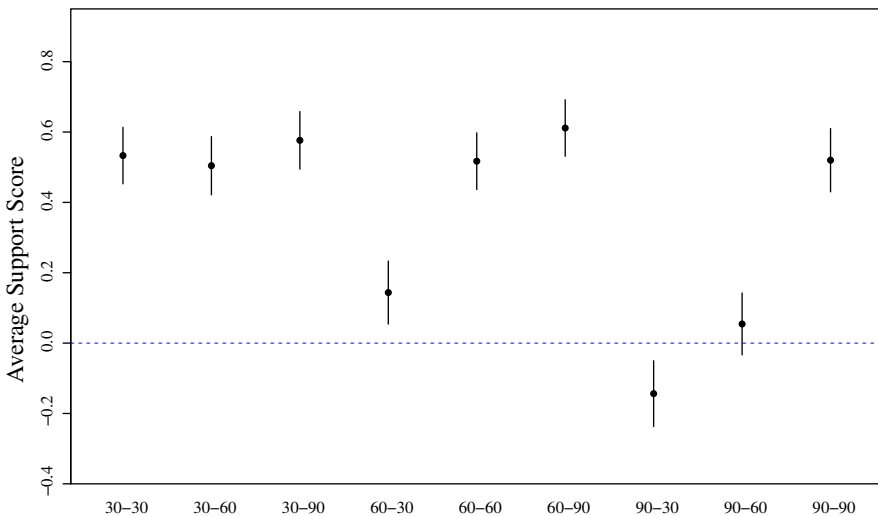


*H2: Americans will be least supportive of outcomes that leave the United States relatively behind, regardless of absolute gains, and the difference in support between equal and unfavorable conditions will be greater than that between favorable and equal conditions.*

*H3: Trade deals in which the United States and another country concede equally will be judged the most fair, while unfavorable trade deals will be judged as substantially less fair than favorable ones.*

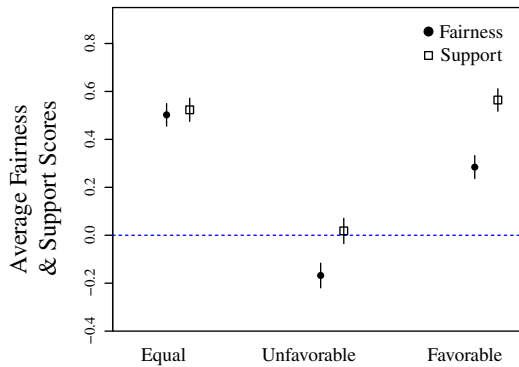
*H4: Fairness will most strongly mediate the difference in support across the unfavorable and the equal conditions since Americans will be particularly aggrieved by injustices done to them.*

We first test whether varying the absolute level of concessions alters levels of support. Figure 1 arrays the various treatments by absolute gains in terms of concessions made to the US from smallest to largest. If absolute gains drive American preferences, then we should see a monotonic increase as we move from left to right, marked by a series of steps. However, no such trend is evident.



*Notes:* Figure displays the average support score by treatment type, measured from -2 to 2, with 95 percent confidence intervals. Higher values represent greater levels of support for the trade agreement. The treatments are arranged from smallest to largest tariff reductions from left to right, where the first number is the percent tariff cut by the US and the second is the percent cut by the other country.

**FIGURE 1.** Support for agreement by treatment



Notes: Figure displays the average fairness and support scores by treatment type, measured from -2 to 2, with 95 percent confidence intervals. Higher values represent greater levels of fairness and support for the trade agreement.

FIGURE 2. *Perceived fairness by treatment*

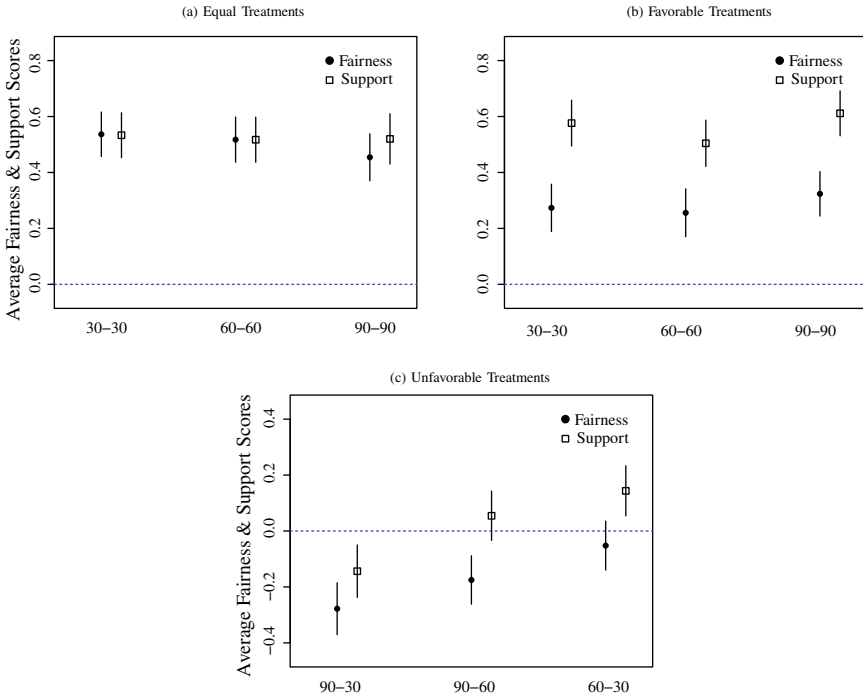
What becomes clear, in contrast, is the importance of relative concessions. Comparing support for the equal, favorable, and unfavorable categories of treatments, grouped together in Figure 2, we see that Americans are equally content with equally balanced agreements and favorable ones. Fifty-one percent of the respondents support the equal agreement and the difference between the percent of respondents supporting the equal and favorable agreements is statistically insignificant (1 percent difference,  $p = 0.47$ ),<sup>32</sup> showing that the public is equally willing to support an agreement where the concessions are equal or favorable to the US.<sup>33</sup> Furthermore, we find that as long as things are balanced or favorable, Americans are content and indifferent to gaining relatively more. As panel (b) shows in Figure 3, which regroups the nine conditions into our three broader categories, support does not increase as relative gains increase.<sup>34</sup>

However, we find that support for the unfavorable trade deals is at least twenty percentage points lower than support for the favorable agreements and the equal agreements ( $p < 0.01$ ), consistent with our expectations. Panel (c) of Figure 3 shows that the least favored outcome is the one in which the ratio of concessions is most skewed toward the other's advantage (90/30). Americans, like human beings in general, are particularly sensitive to injustices done to them.

32. In this context, we count people as supporting the agreement if they “somewhat” or “very strongly” supported it.

33. Comparing the overall support score, the difference between the equal and favorable treatments is also insignificant ( $-0.04$ ,  $p = 0.32$ ).

34. The difference in panel (b) that comes closest to being significant is between the 60/90 and 30/60, but it is only 0.07 ( $p = 0.34$ ).



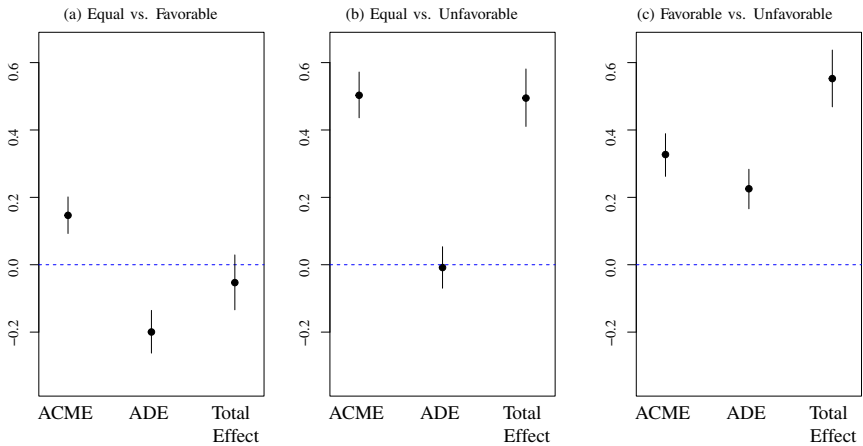
Notes: Figure displays the average fairness and support scores by treatment type, measured from -2 to 2, with 95 percent confidence intervals. Higher values represent greater levels of fairness and support. In panel (a) none of the differences in fairness or support approach statistical significance. In panel (c) the drop in perceived fairness from the 60/30 treatment to the 90/30 treatment is -0.23 ( $p < 0.01$ ) and the drop in the fairness score from 90/60 to 60/30 is 0.12 ( $p = 0.10$ ). None of the differences among the favorable treatments in panel (b) approach significance. The support difference between the 30/60 and 60/90 comes closest at 0.11 ( $p = 0.13$ ) and the fairness difference between the 60/90 and 30/60 comes the closest at -0.07 ( $p = 0.34$ ).

FIGURE 3. Support by treatment

Turning to an evaluation of what respondents regard as fair, Figure 2 presents the average fairness scores across treatments, which shows that the equal deal is viewed as the fairest, consistent with H3. This indicates that fairness judgments are not merely a function of interests; they indicate some degree of impartiality (although this does not translate perfectly into greater support, a point we will explore).

However, our respondents are not entirely impartial in that disadvantageous outcomes are judged as substantially less fair than advantageous ones. Panel (c) of Figure 3 shows that respondents view it as especially unfair when they fall furthest behind in the 90/30 treatment. The drop in perceived fairness from the 60/30 treatment to the 90/30 treatment is -0.23 ( $p < 0.01$ ) and the drop in the fairness score from 90/60 to 60/30 is -0.12 ( $p = 0.10$ ), showing that the public believes it is increasingly unfair when they give up more than the other country. This is in contrast to the

stable perceptions of fairness across all favorable trade deals, which are shown in panel (b) of Figure 3, where there are no significant differences in perceived fairness.<sup>35</sup>



Notes: Figure plots the Average Causal Mediation Effects (ACME), the Average Direct Effects (ADE) and Total Effects from a series of mediation models in which the effect of each treatment on support for the trade agreement is mediated through perceived fairness of the agreement. Both support for the agreement and fairness of the agreement are measured on a five-point scale, where higher values represent more support and higher perceived fairness. Analysis was conducted using the mediation package by Tingley et al. 2014, and includes controls for the following pretreatment covariates: age, education, income, political party, gender, national attachment, and cooperative internationalism.

FIGURE 4. Mediation analysis

We now shift to an examination of whether perceptions of fairness, partial as they are, play an important role in shaping support for trade agreements. To do so we employ mediation analysis.<sup>36</sup> In Figure 4, we display the average causal mediation effect (ACME) of fairness, average direct effect of the treatment (ADE), and the total effect on support for the trade agreement. Hypothesis 4 expects that fairness will most strongly mediate the difference in support between the unfavorable and equal trade outcomes since Americans are predicted to be most bothered by injustices done to them. The mediation effect of fairness is responsible for the entire increase in support for the agreement that results from shifting from an unfavorable to an equal trade agreement (ACME: 0.51,  $p < 0.01$ ), as panel (b) shows. Americans prefer equal outcomes over disadvantageous ones because they judge them to be fairer.

In contrast, when shifting from the favorable to the equal trade agreement, the equal agreement has a negative direct effect on support for the trade agreement, but the mediated effect of fairness pushes in the opposite direction. Therefore,

35. The difference between favorable treatments that comes closest to being significant is the support score for 30/60 to 60/90, which is 0.11 ( $p = 0.13$ ).

36. Imai et al. 2011. A discussion of the limitations and assumptions of the mediation analysis is included in the appendix.

fairness serves to restrain self-interest to the extent that we end up with roughly similar levels of support for equal and favorable outcomes. However, this mediated effect is considerably smaller (about 40 percent) because fairness perceptions are asymmetric in nature. Respondents are less bothered when Americans come out ahead than when others do. We see this most clearly when estimating the mediating effect of fairness on support for favorable versus unfavorable outcomes in panel (c). Fairness is responsible for about 60 percent of the increase in approval that results from moving from an unfavorable to a favorable agreement (ACME: 0.33,  $p < 0.01$ ). This indicates a degree of, but not complete impartiality, exactly what we would expect.

Readers might be skeptical of these mediation results because perceptions of fairness might be endogenous to support for agreements as a result of consistency seeking by respondents to feel morally good about whatever choices they make. In other words, individual biases affect our mediation estimates. However, such bias, which we admit, actually makes the results that much more striking. If such bias were driving respondents' fairness evaluations, we would not observe differences in mediation patterns across the three comparisons since fairness would always push in the same direction as the treatments, justifying and therefore equally mediating whatever respondents prefer. Yet this is not the case at all. Instead, despite the likely tendency for respondents to find fairer that which they support, we still see substantially different mediation patterns in comparing support levels among these three groups of treatments—patterns consistent with our argument.

Skeptics might also wonder whether our results are consistent with an entirely egoistic conception of reciprocity with no need for recourse to “social preferences.” Under this logic, individuals generally prefer fairer, equal outcomes because they lend themselves to more stable long-term outcomes that benefit both sides. In other words, fairer is smarter. While we cannot entirely rule out such an account, which deserves further study, we believe that it is unlikely that it would generate the same heterogeneous treatment effects and fairness mediation dynamics observed in our data. However, we are mindful of such concerns, so we also report in the appendix the effects of an individual-level measure of fairness. We find the equal treatment has a greater effect on support for trade agreements among those who define fairness as equality. This heterogeneous treatment effect contributes to our conclusion that a sense of fairness is an important mediator in explaining support for objectively fairer outcomes, since this individual-level variable suffers from less possibility of endogeneity and bias.

## Trade Balance Study

Equality, while the most common conception of fairness, is not the only one. Another is equity, judged in terms of whether outputs reflect inputs. Are we fairly compensated given what we contributed to some collective process? In a business setting, this is generally reflected in whether an employee's compensation is commensurate

with her productivity. Equity concerns explain human beings' desire to punish free riding. We do not want to split the pie evenly if it means that those who did not contribute to the cooking get to eat it.<sup>37</sup> Most inequity aversion models are in this sense infelicitously named. They should be called inequality aversion. In traditional ultimatum games, equity is irrelevant since neither side contributed anything to the creation of the pot to allocate.

When it comes to trade, equity and equality conceptions typically converge in the notions of "fair play" and a "level playing field." Consider the ideal type of a free trade area. Both sides completely remove their trade barriers so that all products are treated equally. The gains from trade will exhibit fairness as equity in that they accrue to those who produce most efficiently. Complaints about an unfair lack of access to foreign markets could therefore indicate a concern for both equality and equity. If the United States has lowered tariffs and this has gone unreciprocated by others (fairness as equality), American producers will not reap the gains from trade that they deserve as a function of their hard work (fairness as equity). The same logic applies to nontariff barriers as well. For instance, subsidies tilt the playing field (fairness as equality), thereby leading to imbalances in trade that do not reflect efficient production (fairness as equity).

However, fairness as equality and equity are sometimes in tension. Seen in terms of equity, one might rightly and justly expect winners or losers in any competitive endeavor based on different levels of effort. So long as the playing field was equal, that is fair. However, this might result in asymmetric gains that by a different benchmark for equality—for instance, gains in jobs or the trade balance—might be perceived as unfair. In terms of trade, the possibility that equity conceptions of fairness might influence attitudes raises the question: are Americans willing to accept unequal outcomes in terms of the tangible benefits of trade if they reflect relative effort, even when the United States comes out relatively behind?

There are reasons to believe that equity will not affect trade attitudes in such an impartial manner. DeScioli and colleagues randomly assigned survey respondents into one of two roles: typists who transcribed three paragraphs and checkers who proofread one of the three paragraphs.<sup>38</sup> The typist was then given the choice of dividing earnings for the task equally, based on the equality conception of fairness, or based on equity, in which the typist received 75 percent of the money. Typists typically judged the equity option as more fair, whereas checkers judged equality to be fairer. In other words, there is a self-serving aspect in weighing equality against equity.

Based on our asymmetric fairness argument, we expect that equity will matter to individuals only when they do not come out behind in relative terms. In other words, equity cannot compensate for the perceived unfairness of relative losses to which they are particularly averse. However, individuals might be particularly

37. Adams 1965; Anderson and Patterson 2008; Deutsch 1975.

38. DeScioli et al. 2014.

supportive of favorable outcomes when Americans are more deserving and of egalitarian outcomes when both sides contribute equally to the outcome. As long as Americans are not relatively disadvantaged, they can evaluate outcomes impartially.

We test the effect of equity using a second study, which we refer to as the Trade Balance Study. The Trade Balance Study was launched two weeks after the first, and included an equity manipulation in addition to equality manipulations. The second study concerns a trade agreement the US was already a party to, where the agreement led to either an equal, unfavorable, or favorable balance of trade between the US and the other country. This design also allows us to assess whether evaluations of fairness are consistent across different aspects of trade agreements, comparing perceptions of tariff concessions from the first study to trade balances in the second.

In the Trade Balance Study, the randomization of the trade flows was described such that the “agreement has led to a 2 to 1 trade imbalance favoring the United States, in which the United States exports twice as much as it imports from the other country,” or it has led to a “2 to 1 trade imbalance favoring the other country,” or it has led to “a relative balance in trade between the two countries.” We refer to the first as “favorable,” the second as “unfavorable,” and the last as “balanced.” In this manner, the balanced, favorable, and unfavorable treatments are conceptually consistent with those tested in the first study, but are tied to the resulting trade flows between the countries, as opposed to the concessions made during the negotiations.

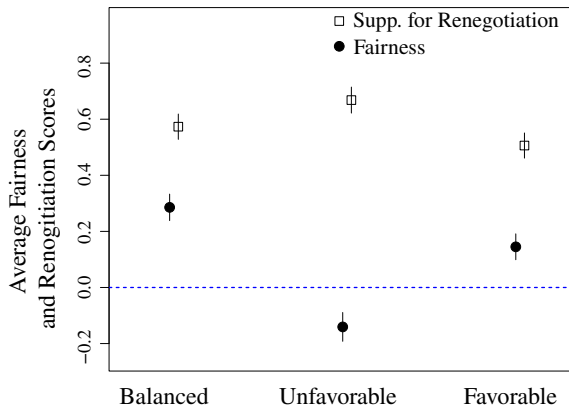
We introduce an equity element as well. To test the effect of equity we might have experimentally manipulated variables such as foreign subsidization of domestic industries or other nontariff barriers to trade. However, since equity is about the work put in and the relative output, we instead utilized a more intuitive manipulation, one based on relative worker productivity.<sup>39</sup> The three potential productivity treatments specified that either “American workers are twice as productive as the workers in the other country,” or “workers in the other country are twice as productive as workers in the United States,” or “workers in both countries are about equally productive.” Work effort and contribution are central to equity considerations of fairness, and lend themselves naturally to a trade scenario.

While the full treatment conditions for the second study are included in the appendix, an example of the wording of one of the treatment combinations, with a favorable trade balance and productive American workers is as follows:

The United States is part of a free trade agreement in which the United States and the other country reduced its tariffs—that is, the tax charged on foreign goods and services when they cross borders. The agreement has led to a 2 to 1 trade imbalance favoring the United States in which the United States exports twice as much as it imports from the other country. Studies show that American workers are twice as productive as the workers in the other country.

39. We evaluate how respondents interpreted the productivity treatment in the appendix.

After reading about the trade agreement, the second study then asked respondents whether they would support *renegotiating* the agreement. In each case respondents could select from “strongly support” to “strongly oppose” on a five-point scale. Respondents were also asked to evaluate their perceived fairness of the trade agreement, as they were in the first study.



*Notes:* Figure displays the average fairness and support for renegotiation scores by treatment type, measured from -2 to 2, with 95 percent confidence intervals. Higher values represent greater levels of fairness and more support for renegotiating the trade agreement. The renegotiation score differences are: balanced versus unfavorable 0.09 ( $p = 0.02$ ), balanced versus favorable 0.07 ( $p = 0.08$ ), and favorable versus unfavorable 0.16 ( $p < 0.01$ ). The differences for all fairness score comparisons are significant ( $p < 0.01$ ).

**FIGURE 5.** *Perceived fairness by trade flows*

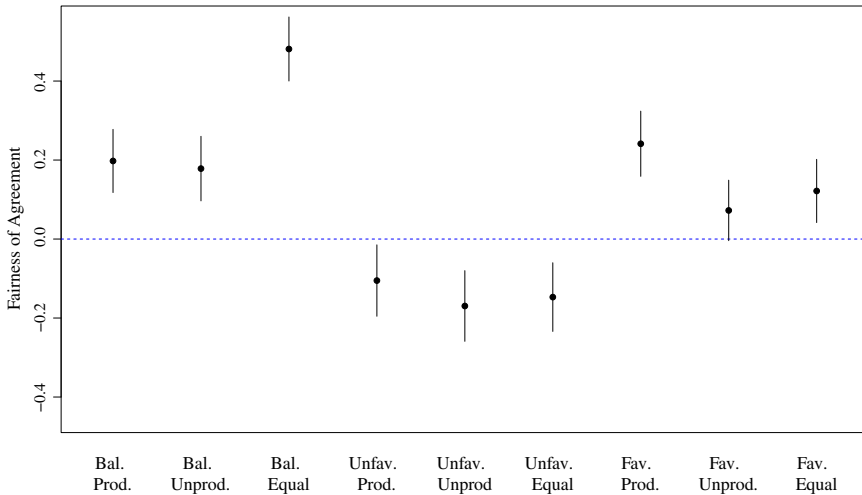
Our expectations for equality are the same as in the previous study, to which we add an additional expectation in light of the inclusion of equity.

*H5: Equitable outcomes in which productivity and relative gains are in line with one another will be perceived as more fair and generate less support for renegotiation only if the United States is not running a trade deficit.*

We begin our analysis of the Trade Balance Study with an examination of the main effects of the balanced, favorable, and unfavorable trade flow treatments. This allows us to assess whether the public’s perceptions of fairness are consistent across the two studies. **Figure 5** displays the average fairness and renegotiation scores for each treatment, showing that trade agreements that result in balanced trade flows are viewed as significantly fairer than both favorable and unfavorable trade flows. The change in the fairness score for moving from the unfavorable treatment to the equal treatment is 0.43 ( $p < 0.01$ ), which corresponds to 13 percent more respondents believing the agreement is fair ( $p < 0.01$ ). Perceptions of fairness are once again significantly higher when the trade flows are equally balanced, as opposed to favorable, with



the fairness score being 0.14 higher in the equal treatment ( $p < 0.01$ ) and 7 percent more of the respondents believing the trade agreement is fair ( $p < 0.01$ ).



Notes: Figure displays the average fairness score by treatment type, measured from -2 to 2, with 95 percent confidence intervals. Higher values represent greater levels of fairness for the trade agreement.

FIGURE 6. Fairness of agreement by equity and equality

When it comes to support for renegotiation, which we interpret as dissatisfaction with the current agreement, there is only a marginal difference in the support score between the balanced and favorable conditions (0.07,  $p = 0.08$ ) and the percent change in respondents supporting renegotiation is not significantly different between the two ( $p = 0.21$ ). Overall the treatment effects on support for renegotiating the agreement are substantively smaller than in our earlier experiment. The direction of our effects, however, is in keeping with our asymmetric fairness argument's expectations. There is a significant change when moving from the balanced to the unfavorable condition, with 6 percent fewer respondents supporting renegotiation in the unfavorable condition ( $p = 0.01$ ). This analysis shows that the relative balance of trade shapes support for renegotiating the agreement and that respondents once again have an asymmetric reaction, responding more strongly to the unfavorable condition than to the favorable condition. Even as we shift the scenario from one about the distribution of concessions to the distribution of benefits, Americans use the same standard for evaluating trade agreements.

We next analyze the fairness judgments across all combinations of productivity and trade flow treatments. The results appear in Figure 6. When considering the unfavorable balances—shown in the middle of the figure—a lack of American productivity cannot compensate for the perceived unfairness of the relative loss. Once respondents feel as though they are falling behind, they are no longer worried about equity and appear to focus on only whether they are doing worse than others.

However, equity considerations are important in the favorable and balanced outcomes. When we shift back to analyzing equity concerns in the favorable condition—shown in the right portion of [Figure 6](#)—respondents viewed the equitable condition—where US workers are more productive and there is a favorable trade flow—as fairer than the inequitable combinations (0.17,  $p = 0.01$  and 0.12,  $p = 0.09$ ). The outcome judged most fair is the one in which the trade flows are equally balanced and workers are equally productive. When inputs and outputs are commensurate, respondents judged it to be significantly fairer than the other two balanced treatments (0.28,  $p < 0.01$  and 0.30,  $p < 0.01$ ), and also the treatment when the United States is more productive and gains more. Americans find fairest outcomes that are both equal and equitable, thus maximizing both conceptions of fairness.

Nevertheless, in this experiment, distinctions in the fairness of outcomes based on equity do not seem to translate into differences in overall levels of support. For instance, respondents in conditions in which the trade balance is equal are no less likely to support renegotiation when productivity is equal than when one side is more efficient.<sup>40</sup> This suggests that when it comes to drawing overall conclusions about trade agreements, fairness as equality is more important than fairness as equity. When we aggregate fairness scores into three groups of treatments, as we did in the previous experiment, we find the same pattern of fairness judgments: equal outcomes are judged the fairest, and advantageous ones are seen as fairer than disadvantageous.

## Conclusion

Our results show that fairness matters for how the mass public in the United States thinks about trade. In our first study, we find that Americans value equality in international trade and they prefer balanced trade agreements in which the distribution of concessions is equal. Not only does the public view such agreements as being fairer than favorable and unfavorable agreements, but this concern for fairness leads to increased support for balanced trade agreements. In our second study, we show that the public is also concerned about equity in trade, when inputs match outputs, although within limits imposed by their commitment to inequity aversion. If Americans come out relatively behind, equity does not compensate for the perceived lack of fairness. Only when Americans come out ahead or equal do they rate fairer outcomes in which the trade balance reflects productivity.

Importantly, fairness judgments do not simply follow the flag, but they are not entirely impartial either. Americans judge balanced trade deals as more fair than favorable ones, and they even support many balanced trade outcomes as much or more than those in which the United States comes out ahead both in relative or absolute terms. Americans are, however, more inclined to see injustices done to them than

40. See appendix for further analysis.

to others. They consistently rate unfavorable trade deals as the least fair, even if workers in the United States are not as productive as those in the country doing better, and they have lower levels of support for these trade deals. Nevertheless, they significantly prefer trade agreements in which each side gains equally, especially when the outputs are commensurate with inputs. In this sense, we can say that they are truly committed to everyone getting a fair share.

This asymmetric evaluation of trade is different than the traditional way in which international relations scholars handle relative distributions in state interactions—generally fixated on whether states pursue relative gains. As a whole it does not appear that the American public prefers relative or absolute gains but they are very concerned about relative losses. A focus on fairness, lacking in previous scholarship, brings this story into light and highlights the asymmetric lens through which fairness is evaluated.

Our results also add further support to a growing body of literature that argues attitudes toward trade are not just about economic interest.<sup>41</sup> Although our studies show that Americans exhibit significant concern for falling behind in trade agreements, they are also willing to forgo both relative and absolute gains in favor of trade deals that they believe are fairer. The importance of fairness in shaping attitudes toward trade helps explain why politicians repeatedly and loudly use fairness as the rallying cry for trade policies. Although Americans show a strong willingness to support equal trade agreements, they also react strongly in opposition of unfavorable trade deals that they believe are unfair.

## Data Availability Statement

Replication files for this research note may be found at <<https://doi.org/10.7910/DVN/THNBPD>>.

## Supplementary Material

Supplementary material for this research note is available at <<https://doi.org/10.1017/S0020818321000084>>.

## References

- Adams, J. Stacy. 1965. Inequity in Social Exchange. In *Advances in Experimental Social Psychology*, vol. 2, edited by Leonard Berkowitz, 267–99. Elsevier.
- Anderson, William D., and Miles L. Patterson. 2008. Effects of Social Value Orientations on Fairness Judgments. *The Journal of Social Psychology* 148 (2):223–46.

41. Ehrlich 2010, 2018; Margalit 2012; Mutz and Kim 2017

- Axelrod, Robert, and William Donald Hamilton. 1981. The Evolution of Cooperation. *Science* 211 (4489): 1390–96.
- Babcock, Linda, and George Loewenstein. 1997. Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives* 11 (1):109–26.
- Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. Biased Judgments of Fairness in Bargaining. *The American Economic Review* 85 (5):1337–43.
- Bagwell, Kyle, and Robert W. Staiger. 1999. An Economic Theory of GATT. *American Economic Review* 89 (1):215–48.
- Bechtel, Michael M., Federica Genovese, and Kenneth F. Scheve. 2019. Interests, Norms and Support for the Provision of Global Public Goods: The Case of Climate Co-Operation. *British Journal of Political Science* 49 (4):1333–55.
- Bechtel, Michael M., and Roman Liesch. 2020. Reforms and Redistribution: Disentangling the Egoistic and Sociotropic Origins of Voter Preferences. *Public Opinion Quarterly* 84 (1):1–23.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10 (1):122–42.
- Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. Separating the Shirkers from the Workers? *American Journal of Political Science* 58 (3):739–53.
- Blount, Sally. 1995. When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences. *Organizational Behavior and Human Decision Processes* 63 (2):131–44.
- Brutger, Ryan, and Joshua D. Kertzer. 2018. A Dispositional Theory of Reputation Costs. *International Organization* 72 (3):693–724.
- DeScioli, Peter, Maxim Massenkoff, Alex Shaw, Michael Bang Petersen, and Robert Kurzban. 2014. Equity or Equality? Moral Judgments Follow the Money. *Proceedings of the Royal Society of London B: Biological Sciences* 281 (1797):20142112.
- Deutsch, Morton. 1975. Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? *Journal of Social Issues* 31 (3):137–49.
- Ehrlich, Sean. 2018. *The Politics of Fair Trade: Moving Beyond Free Trade and Protection*. Oxford University Press.
- Ehrlich, Sean D. 2010. The Fair Trade Challenge to Embedded Liberalism. *International Studies Quarterly* 54 (4):1013–33.
- Fehr, Ernst, and Urs Fischbacher. 2002. Why Social Preferences Matter—The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives. *The Economic Journal* 112 (478):C1–C33.
- Fehr, Ernst, and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114 (3):817–68.
- Finalyson, Jock A., and Mark W. Zacher. 1981. The GATT and the Regulation of Trade Barrier: Regime Dynamics and Functions. *International Organization* 35 (4):561–602.
- Fiske, Alan Page. 1991. *Structures of Social Life: The Four Elementary Forms of Human Relations: Communal Sharing, Authority Ranking, Equality Matching, Market Pricing*. Free Press.
- Frank, Morgan R., Nick Obradovich, Lijun Sun, Wei Lee Woon, Brad L. LeVeck, and Iyad Rahwan. 2018. Detecting Reciprocity at a Global Scale. *Science Advances* 4 (1):eaao5348.
- Gottfried, Matthew S., and Robert F. Trager. 2016. A Preference for War: How Fairness and Rhetoric Influence Leadership Incentives in Crises. *International Studies Quarterly* 60 (2):243–57.
- Grieco, Joseph M. 1988. Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism. *International Organization* 42 (3):485–507.
- Guisinger, Alexandra. 2017. *American Opinion on Trade: Preferences Without Politics*. Oxford University Press.
- Herrmann, Richard K., Philip E. Tetlock, and Matthew N. Diascro. 2001. How Americans Think About Trade: Reconciling Conflicts Among Money, Power, and Principles. *International Studies Quarterly* 45 (2):191–218.
- Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. Unpacking the Black Box of Causality: Learning About Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review* 105 (4):765–89.

- Keohane, Robert O. 1986. Reciprocity in International Relations. *International Organization* 40 (1):1–27.
- Kertzer, Joshua D., and Ryan Brutger. 2016. Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory. *American Journal of Political Science* 60 (1):234–49.
- Kertzer, Joshua D., and Brian C. Rathbun. 2015. Fair Is Fair: Social Preferences and Reciprocity in International Politics. *World Politics* 67 (4):613–55.
- Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. 1989. Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology* 57 (3):426–41.
- Lü, Xiaobo, Kenneth Scheve, and Matthew J. Slaughter. 2012. Inequity Aversion and the International Distribution of Trade Protection. *American Journal of Political Science* 56 (3):638–54.
- Mansfield, Edward D., and Diana C. Mutz. 2009. Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety. *International Organization* 63 (3):425–57.
- Margalit, Yotam. 2012. Lost in Globalization: International Economic Integration and the Sources of Popular Discontent. *International Studies Quarterly* 56 (3):484–500.
- Mutz, Diana C., and Eunji Kim. 2017. How Ingroup Favoritism Affects Trade Preferences. *International Organization* 71 (4):827–50.
- Pinker, Steven. 2012. *The Better Angels of Our Nature: Why Violence Has Declined*. Penguin.
- Powell, Robert. 1991. Absolute and Relative Gains in International Relations Theory. *American Political Science Review* 85 (4):1303–20.
- Rabin, Matthew. 1993. Incorporating Fairness into Game Theory and Economics. *The American Economic Review* 83 (5):1281–302.
- Rabin, Matthew. 1998. Psychology and Economics. *Journal of Economic Literature* 36 (1):11–46.
- Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. Mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software* 59 (5):1–38.

## Authors

**Ryan Brutger** is Assistant Professor in the Travers Department of Political Science at University of California, Berkeley. He can be reached at [brutger@berkeley.edu](mailto:brutger@berkeley.edu).

**Brian Rathbun** is Professor at University of Southern California's School of International Relations. He can be reached at [brathbun@usc.edu](mailto:brathbun@usc.edu).

## Acknowledgments

We thank Alex Baturo, Eunji Kim, Diana Mutz, Nina Rathbun, Peter Rosendorff, Mike Tomz, Jessica Weeks, the audiences at ISA, IPES, and UCSD, and others who have provided feedback on earlier versions of this research.

## Key Words

Trade; public opinion; fairness; equality; equity

Date received: December 3, 2019; Date accepted: July 15, 2020