Collective risk social dilemma and the consequences of the US withdrawal from international climate negotiations

Oleg Smirnov
Stony Brook University, USA

Abstract
International climate negotiations represent an interesting theoretical problem, which can be analyzed as a collective risk social dilemma as well as an n-person bargaining model. The problem is made more complicated by politics due to the differences between: (1) total and per capita emissions; and (2) present-day and cumulative emissions. Here, we use a game theoretic approach in conjunction with the literature on effort-sharing approaches to study a model of climate negotiations based on empirical emissions data. We introduce a ‘fair equilibrium’ bargaining solution and examine the consequences of the United States’ withdrawal from the Paris Agreement. Our results suggest that the collective goal can still be reached but that this requires additional greenhouse gas emissions cuts from other countries, notably, China and India. Given the history of climate negotiations, it is unclear if these countries will have sufficient political will to accept the additional costs created by the US defection.

Keywords
Climate change mitigation; collective risk social dilemma; effort-sharing approaches; fair equilibrium; Paris Agreement

Corresponding author:
Oleg Smirnov, Department of Political Science, Stony Brook University, SBS-7, Stony Brook, New York, USA.
Email: oleg.smirnov@stonybrook.edu
1. Introduction

The Paris Agreement approved on December 12, 2015 by representatives of 196 countries (United Nations Framework Convention on Climate Change Conference of the Parties, 2015) signified an important step forward to keeping global warming under 2°C above pre-industrial levels—a critical temperature increase threshold identified by the Intergovernmental Panel on Climate Change (IPCC) (Intergovernmental Panel on Climate Change, 2013). The agreement paves the way for leaders to make the emissions-reduction pledges that would be necessary for the world to meet the collective goal—that depending on their nations being willing to stay committed to matching contributions with their promises. However, on June 1, 2017, the United States President Donald Trump announced that the US would withdraw from the Paris Agreement (Bordoff, 2017; Liptak and Merica, 2017; Pickering et al., 2018; Urpelainen and Van de Graaf, 2018). The withdrawal is not effective until November 2020 according to Article 28 of the Paris Agreement (United Nations Framework Convention on Climate Change Conference of the Parties, 2015). Exploring the consequences of the US exit is one of the most important research questions today (Dai et al., 2018; Nong and Siriwardana, 2018).

What are the prospects of international climate negotiations if the US does not contribute to the global emissions-reduction goal? Will such defection derail climate change mitigation efforts? What would other nations have to do in order to achieve the collective goal despite the defection by the US?

Here we address these questions using a game theoretic model. Game theory allows us to model strategic interaction between multiple players. The players are assumed to be rational, that is, maximizing their payoffs in the game. In equilibrium, no player has an incentive to deviate from the currently chosen strategy (Nash, 1950b). In game theoretic terms, climate change mitigation has been recently seen as a collective-risk social dilemma (CRSD) (Milinski et al., 2008; Tavoni et al., 2011): the players benefit when the public good is provided but no player has an incentive to pay the cost associated with that provision. However, unlike in the famous mode of the ‘tragedy of the commons’ (Hardin, 1968), mutual defection is not the only equilibrium in the game. In the CRSD model, the players make contributions to reach a certain threshold—a common goal—and, if they do so, avoid a high probability of losing all their resources. In a cooperative equilibrium, the sum of players’ contributions is equal to the threshold and no player has an incentive to defect and, thereby, to face a risk of total loss (Bynum et al., 2016; Erev and Rapoport, 1990; Offerman et al., 1998).

In the context of climate change mitigation, national contributions correspond to cuts in greenhouse gas (GHG) emissions. Such contributions are potentially costly since they may cause higher energy prices, slower economic growth, and require major immediate investments in alternative energy solutions (Shackleton, 2009). Unlike in the linear public good games, however, the benefit of the individual contributions can only be obtained if the sum of individual contributions is large enough to prevent abrupt climate change. If the threshold value is not reached then all individual contributions represent a wasted effort.
According to the most recent IPCC report, global emissions in 2017 were 53.5 gigatonnes of carbon dioxide (GtCO₂). To limit global warming to 2°C (temperature increase relative to pre-industrial levels), the emissions must be decreased by about 25% by 2030 (Intergovernmental Panel on Climate Change, 2018). Exceeding the threshold value set by climate scientists is dangerous due a higher likelihood of extreme weather events and deviation-amplifying mutual causal processes, or ‘climate feedbacks’ (Intergovernmental Panel on Climate Change, 2013). All countries in the world prefer to avoid abrupt climate change; at the same time, they also want to minimize the costs associated with individual national contributions. Therefore, the collective risk social dilemma is a good model capturing the problem.

Previous applications of the CRSD in the context of climate change mitigation are based on highly stylized models of strategic interaction of hypothetical players; while these models do make important theoretical insights, their policy relevance is limited since the modeling parameters are not based on empirical data (Madani, 2013). Another possible limitation of the game theoretic analysis of a generic CRSD is due to the fact that there is a large number of cooperative equilibria. As a result, it is not clear which of the equilibria are more feasible than others. A salient cooperative equilibrium (Van Huyck et al., 1991) in which all players contribute equally to the common goal is not realistic since we cannot expect all countries to contribute equally to the global emissions reduction target. Similarly, the classic Nash bargaining solution (Nash, 1950a) and the Kalai–Smorodinsky alternative (Kalai and Smorodinsky, 1975) are not appropriate solutions since they only rely on the present-day national disagreement values and do not take into account the reality of the situation: namely, (1) per capita emissions, and (2) historical responsibility of the developed nations—that is, the required fairness considerations.

In the present paper, we make an effort to overcome these problems by: (1) offering a version of the CRSD based on empirical emissions data and future projections; and (2) proposing a unique bargaining solution based on (any) given fair allocation rule. The proposed solution is then applied to examine the impact of the US withdrawal from the international climate negotiations.

2. Model

Consistent with the bargaining model of international climate negotiations by Smead et al. (2014), each country has a size $s_i$ equal to the country’s emissions under a business-as-usual scenario. The country size is analogous to the endowment value in the standard CRSD (Milinski et al., 2008).

Each country then makes a demand $0 \leq d_i \leq 1$, corresponding to a proportion of $s_i$. In a CRSD, this is analogous to the proportion of the endowment that the players do not contribute to the public goal.

Each country generally prefers to make a higher demand; however, if the sum of the countries’ demands, weighted by their respective sizes, is greater than the global emissions threshold $T \sum s_i$ ($0 < T < 1$) then player $i$’s payoff $\pi_i$ is only a fraction of the demand, $d_i \delta$, where $0 < \delta < 1$ is the disagreement value:
The payoff is reminiscent of the classical CRSD, in which individual contributions must reach the threshold value. Compare this to the present model, in which individual non-contributions (demands) must not exceed the threshold. The ability to use the bargaining framework is the main reason why we are adopting the non-contributions (demands) approach instead of the familiar contributions model.

Notice that if the disagreement value $\delta$ is small then there is a strong incentive to moderate GHG emissions since the potential loss is large. On the other hand, if $\delta \to 1$ then there are no negative consequences for exceeding the global emissions threshold and cooperation is virtually impossible (as well as not rational). This is yet another feature of the standard CRSD, in which the disagreement value is virtually the same as the risk of a loss exogenous parameter.

The parameter $T$ represents a value that should not be exceeded by the sum of individual non-contributions (similarly, in a standard CRSD, the threshold value must be reached by the sum of contributions). Here, a large parameter $T$ indicates that meeting the collective goal is relatively easy. Conversely, a smaller $T$ implies a bigger challenge for the players. For instance, keeping the global temperature increase under $1.5^\circ C$ corresponds to a lower $T$, relative to the $T$ associated with the $2^\circ C$ goal.

### 3. Fair equilibrium

It is well-known that in the CRSD, there are two kinds of equilibrium outcomes (Erev and Rapoport, 1990; Heitzig et al., 2011; Milinski et al., 2008; Offerman et al., 1998; Smead et al., 2014):

1. In one equilibrium, everyone defects. No player has a capacity to reach the common goal unilaterally. Therefore, everyone is stuck defecting.
2. In the other equilibrium, everyone contributes just enough to reach the common goal, and no player has an incentive to either increase or decrease their contribution. Each player obtains a payoff greater or equal to their disagreement value.

While the first type of equilibrium is unique, the second one is not since there is an infinite number of possible combinations of contributions the sum of which is equal to the target value. Even if the contributions are restricted to the integer values, the number of cooperative equilibria is still very large.

To increase policy-relevance of a game theoretic solution, we should make an effort to reduce the number of predictions. Ideally, the number of equilibria should be equal to one. A number of bargaining solutions have been proposed, for example, the Nash bargaining (Nash, 1950a) and Kalai–Smorodinsky solutions (Kalai and Smorodinsky, 1975). These classical solutions can be used to analyze international climate negotiations if the latter are modeled as a bargaining problem. The
Nash bargaining solution, for instance, maximizes the total surplus that the players obtain relative to their disagreement values (Nash, 1950a). On the other hand, the Kalai–Smorodinsky solution equalizes the players’ surplus and maximizes that value (Kalai and Smorodinsky, 1975).

In the context of climate change mitigation, one possible issue with these approaches is that they potentially ignore the political dimensions of the bargaining problem such as: (1) the difference between total and per capita emissions; and (2) the difference between the present and cumulative emissions. In order to include these variables into the analysis, we need to extend both the model and its equilibrium solution.

To achieve this goal in the context of the CRSD game, we introduce a novel bargaining solution and call it a ‘fair equilibrium’ (not to be confused with the ‘fairness equilibrium’ (Rabin, 1993), in which players reciprocate kindness of their partners).

Definition: A fair equilibrium is a Nash equilibrium (Nash, 1950b) that minimizes the difference between a set of players’ strategies and an exogenous set of strategies associated with the most equitable outcome.

Thus, a fair equilibrium requires an exogenously defined reference point, describing which outcome is fair. In many real-world situations, such guidance does, in fact, exist. In the context of climate change mitigation, there are several alternative reference points based on the alternative effort-sharing approaches, such as common but differentiated convergence (CDC) (Höhne et al., 2006) and equal cumulative per capita (ECPC) (Bode, 2004) allocation rules. In the present paper, we also use the CDC and ECPC approaches because they were chosen as the key effort-sharing approaches in a seminal research article published in the flagship journal on climate change (Meinshausen et al., 2015). With that said, we emphasize that the fair equilibrium solution can be applied for any effort-sharing approach.

According to the CDC approach, both developed and developing countries are expected to converge to the same (‘common’) level of per-capita emissions within the same convergence period. On the other hand, developing countries should begin cutting emissions only after they reach per-capita emissions that exceed the global average by a pre-determined percentage; hence, the convergence is ‘differentiated’ (Höhne et al., 2006). According to the ECPC approach, national contributions are based on historical per-capita emissions within a certain period (Bode, 2004), for example, since 1950 for the ECPC50 allocation (Meinshausen et al., 2015).

Figure 1(a) shows an intuitive example of fair equilibria for a two-player version of the CRSD game. Furthermore, Figure 1(a) illustrates four different cases, for example, four different allocation rules. Since we are not imposing any restrictions, the most equitable outcomes (Φ) may lie above, below, or on the threshold line. The fair equilibrium is a point on the threshold line that is closest to the most equitable outcome, subject to the disagreement values for all players. If the point closest to the most equitable outcome is located below a disagreement value for a given player (e.g., $D_4$ in Figure 1(a)), then the player will make a demand equal to their disagreement value ($D_4^*$) as long as the point lies on the threshold line.
Figure 1(b) illustrates a scenario with a lower global emissions threshold. In this case, the range of cooperative equilibria is smaller. With fewer options, it is potentially more difficult to arrive at an equitable outcome. All else equal, a larger country is more sensitive to the changes in the emissions threshold. However, if according to a fair allocation rule, a country is expected to make a demand below the disagreement value, it will either exit the agreement or make a demand equal to the disagreement value—that is, a demand exceeding its fair allocation. In this case, the other player(s) will have to adjust (i.e., decrease) their demands accordingly.

Formally, to find the fair equilibrium $D^*_i = d^*_i$ in the CRSD model with $N$ players, the threshold parameter $0 < T < 1$, and the disagreement value $0 < \delta < 1$, we need to minimize the Euclidian distance between the set of Nash equilibrium demands $D = [d_i^{NE}]$ and an exogenously given fair allocation demand $\Phi = [\phi_i]$ for all players:

$$D^* = \arg \min_{d_i} \sum_{i=1}^{N} (d_i^{NE} - \phi_i)^2,$$

subject to the following constraints:

$$\sum_{i=1}^{N} d_i^* s_i = T \sum_{i=1}^{N} s_i \quad \text{(the threshold value is reached but not exceeded)},$$

and

$$d_i^* \geq \delta \forall i \in N \quad \text{(the demands are not below the disagreement value)}.$$
Notice that mutual defection can be the fair equilibrium if it is the only Nash equilibrium in the game. More generally, if a game has a unique Nash equilibrium, then the fair equilibrium and the Nash equilibrium are the same. If the player strategies are continuous, the fair equilibrium is also unique (see below). In the absence of exogenous constraints, the fair equilibrium is the same as the most equitable outcome if the latter is a Nash equilibrium. In the presence of constraints—such as the players’ disagreement values or the threshold value in the CRSD model—the fair equilibrium may deviate from the most equitable outcome as we illustrated in Figure 1.

In a CRSD game with multiple Nash equilibria and continuous player strategies, the fair equilibrium is unique. The fair allocation demand \( \Phi = \{d_i\} \forall i \in N \) is a point in the \( N \)-dimensional strategy space where \( N \) is the number of players. The collective risk threshold (expression (3)) is an \( (N-1) \)-dimensional hyperplane \( H \) that divides the strategy space into two parts: (1) the combinations of strategies that satisfy the threshold constraint; and (2) the combinations of strategies that exceed the total demand threshold. In two dimensions \( (N = 2) \), the hyperplane is a line (e.g., \( [A, B] \) in Figure 1). The shortest distance between \( \Phi \) and \( H \) is equal to the radius of a circle such that \( \Phi \) is the center of the circle, and \( H \) is tangent to the circle (e.g., \( [\Phi_4, D_4] \) in Figure 1). In three dimensions, the result is the same except that \( \Phi \) is the center of a sphere and the threshold hyperplane \( H \) is a two-dimensional surface tangent to the sphere. In \( N \) dimensions, \( \Phi \) is the center of an \( N \)-sphere and \( H \) is the tangent hyperplane. The fair equilibrium is a set of strategies at the only point where \( H \) touches the \( N \)-sphere. Therefore, the fair equilibrium is unique.

Positive disagreement values decrease the space where the players’ strategies satisfy the threshold constraint (e.g., the green area to the right of the \( [A, B] \) line in Figure 1). Since the players in the game never make demands below their disagreement values, in equilibrium their demands are equal to the disagreement value. This is equivalent to decreasing the value of the threshold for the other players in the game, who have yet to reach the threshold—that is, changing the hyperplane \( H \) to a new \( H' \). Then the fair equilibrium is a set of strategies at the only point where \( H' \) touches the \( N \)-sphere with the center at \( \Phi \). Thus, the fair equilibrium remains unique.

### 4. Two-player version of the model

For the sake of intuition, consider a two-player version of the game. In this case, the global climate change threshold is defined by:

\[
d_x s_x + d_y s_y = T(s_x + s_y),
\]

where \( 0 < T < 1 \) is the exogenous parameter; \( d_i, i = [x, y] \) is the demand for country \( i \); and \( s_i, i = [x, y] \) is the size of country \( i \). The payoffs to the players are:

\[
\pi_x = \begin{cases} 
  d_x & \text{if} \quad d_x s_x + d_y s_y \leq T(s_x + s_y) \\
  d_x \delta & \text{if} \quad d_x s_x + d_y s_y > T(s_x + s_y)
\end{cases}
\]

(6.1)
where \(0 < \delta < 1\) is the value of disagreement.

Fair equilibrium \(D^* = [d^*_x, d^*_y]\) is a Nash equilibrium that minimizes the difference between the set of players’ strategies and an exogenous set of strategies associated with the most equitable outcome \(\Phi = [\phi_x, \phi_y]\) subject to the disagreement value \(\delta\) and the global threshold:

\[
[d^*_x, d^*_y] = \arg\min((d^\text{NE}_x - \phi_x)^2 + (d^\text{NE}_y - \phi_y)^2). \tag{7}
\]

To find \(D^*\), first we need to rewrite expression (5) as:

\[
d_y = \frac{T(s_x + s_y) - d_xs_x}{s_y}. \tag{8}
\]

The slope of \([A, B]\) is equal to \(-\frac{s_y}{s_x}\). Since \([A, B]\) is tangent to the circle with the center \(\Phi\), the radius of the circle \([\Phi, D^*]\) is perpendicular to the tangent. Therefore, the slope of the line that contains the radius of the circle is \(-\frac{1}{-\frac{s_y}{s_x}} = \frac{s_x}{s_y}\). Thus, the radius lies on the line:

\[
d_y = \frac{s_y}{s_x} d_x + \beta_0, \tag{9}
\]

where \(\beta_0\) is the intercept. Since \(\Phi = [\phi_x, \phi_y]\) lies on the same line and since it is known, we can express the intercept as:

\[
\beta_0 = \phi_y - \frac{s_y}{s_x} \phi_x. \tag{10}
\]

Hence, we can rewrite expression (9) as:

\[
d_y = \frac{s_y}{s_x} d_x + \phi_y - \frac{s_y}{s_x} \phi_x. \tag{11}
\]

The fair equilibrium lies at the intersection of the global threshold (expression (8)) and the line that contains the radius of the circle (expression (11)):

\[
\frac{T(s_x + s_y) - d_xs_x}{s_y} = \frac{s_y}{s_x} d_x + \phi_y - \frac{s_y}{s_x} \phi_x. \tag{12}
\]

Thus, the equilibrium demands are:

\[
d^*_x = \frac{T s_x (s_x + s_y) + \phi_x s_x^2 - \phi_y s_y s_x}{s_x^2 + s_y^2}, \tag{13.1}
\]

and (substituting expression (13.1) in expression (8)):

\[
\pi_y = \begin{cases} 
    d_y & \text{if } d_x s_x + d_y s_y \leq T(s_x + s_y) \\
    d_y \delta & \text{if } d_x s_x + d_y s_y > T(s_x + s_y)
\end{cases} \tag{6.2}
\]


\[ d_y^* = \frac{T s_y (s_x + s_y) + \phi_y s_x^2 - \phi_x s_x s_y}{s_x^2 + s_y^2}. \]  

(13.2)

Given the fair equilibrium, we observe that a lower threshold has a larger (negative) impact on a larger country since:

\[ \frac{\partial d_x^*}{\partial T} = \frac{s_y (s_x + s_y)}{s_x^2 + s_y^2}, \quad \frac{\partial d_y^*}{\partial T} = \frac{s_y (s_x + s_y)}{s_x^2 + s_y^2}. \]  

(14)

Thus, other things held constant, larger countries are more sensitive to changes in the global climate change threshold.

The solution above (expressions (13.1) and (13.2)) is valid only if the demands do not fall below the disagreement value, that is, \( d_x^* \geq \delta \), and \( d_y^* \geq \delta \). If for a given fair allocation approach, a country’s demand is required to decrease below the disagreement value, then its fair equilibrium demand is equal to the disagreement value—whereas the other country has to decrease its demand accordingly (see also Figure 1 for an illustration in the case of \( \Phi_4 \)), for example,

\[ d_x^* = \delta, \]  

(15.1)

and

\[ d_y^* = \frac{T (s_x + s_y) - \delta s_x}{s_y}. \]  

(15.2)

Thus, it follows that if a country’s fair equilibrium demand is already equal to the disagreement value, then any extra cost associated with a lower threshold must be absorbed by the other country, which has yet to reach the disagreement value.

If the fair demands for both countries are below the disagreement value, then the cooperative fair equilibrium is possible only under the knife-edge condition:

\[ d_x^* = d_y^* = \delta = T. \]  

(16)

The result is easy to derive since the cooperative Nash equilibrium exists if \( \delta s_x + \delta s_y = T (s_x + s_y) \), or \( \delta = T \). Otherwise, the only Nash equilibrium in the game is mutual defection, that is, \( d_x^* = d_y^* = 1 \).

5. Empirical application of the model

To examine the impact of the US withdrawal on the strategic behavior of other nations, we construct a game-theoretic model on the basis of available empirical data. We used Python SciPy optimization to obtain all numerical solutions. We also used Mathematica built-in maximization commands to replicate the results (see the online Appendix).

The number of players in the game is \( N = 16 \), representing the 16 top emitters of GHGs (World Resources Institute, 2014) responsible for about 75% of global
emissions in 2010 (Table 1). Country size $s_i$ is based on the Climate Action Tracker 2030 emissions projections at the time of the 2015 Paris Agreement and current national policies (CAT Consortium, 2015). The target value $T$ is calculated as the proportion of the 2030 emissions required for the world to meet the 2°C goal divided by the total 2030 emissions under the current policies (CAT Consortium, 2015; Meinshausen et al., 2015). The fair allocation rules that we use—ECPC50 and CDC—are based on Meinshausen et al. (2015).

Recall that the reference point in the model is defined as a set of national demands, calculated on the basis of a given effort-sharing approach. To calculate the fair allocation demands, we assume that the 2030 emissions under the current policies correspond to $d = 1$ for all countries. Using Meinshausen et al. (2015) calculations, we know the amount of GHG emissions for individual countries for each allocation rule. For example, China’s maximum possible demand is 14.1 GtCO$_2$. According to the CDC approach, China’s fair amount of emissions has to represent a 32% reduction relative to its emissions in 2010:

$$\Phi_{GtCO2,\text{China}} = 10.164 - 10.164 \times 0.32 = 6.912.$$  \hspace{1cm} (5)

Now we can use this number to calculate the fair demand:

$$\Phi_{d,\text{China}} = 
\frac{6.912}{14.1} \approx 0.49.$$  \hspace{1cm} (6)

Similar calculations are made for both fair allocation rules for all countries in Table 1. The values in the ‘Expected national demands’ column of Table 1, therefore, represent the most equitable demands, which the nations should try to achieve subject to the given constraints: global emissions threshold; and national disagreement values.

In the present paper, we do not attempt to derive specific national disagreement values $\delta_i$. We point out that, similar to Smead et al. (2014), there is a single disagreement value for all countries $\delta_i = \delta / n_i$. Theoretically, this limitation does not present a problem since the fair equilibrium can be just as easily found for heterogenous disagreement values. On the other hand, an empirical application of the model with different deltas is much more challenging since it requires modeling and estimating national disagreements for all countries in the model.

Using a homogenous disagreement value for all countries, we are able to explore the whole parameter space by incrementing the delta from 0 to 1 (Figure 2). In addition, we provide specific illustration for $\delta = [0.2, 0.4, 0.5]$ in Table 2.

For each allocation rule (ECPC50 and CDC) we analyze two scenarios: ‘baseline’; and ‘US exit’. In the baseline scenario, national disagreement values are equal to $0 < \delta < 1$. In the US exit scenario, the disagreement value for one country, the United States, is fixed: $\delta_{US} = 1$. Therefore, it will not cut its GHG emissions, making it more challenging for the rest of the world to stay below the global emissions threshold. Fair equilibria under the ECPC50 allocation rule are shown in Figure 2. Specific examples of the fair equilibria for $\delta = [0.2, 0.4, 0.5]$ are described in Table 2. Similar results for the CDC allocation rule are presented in the online Appendix Figure A1 and Appendix Table A1(a).
<table>
<thead>
<tr>
<th>Rank</th>
<th>2010 (World Resources Institute)</th>
<th>2010 (Climate Action Tracker (CAT))</th>
<th>2030 (CAT)</th>
<th>ECPC50</th>
<th>CDC</th>
<th>ECPC50</th>
<th>CDC</th>
<th>ECPC50/2030 emissions</th>
<th>CDC/2030 emissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>China 9.765</td>
<td>10.164</td>
<td>14.100</td>
<td>-4</td>
<td>-32</td>
<td>9.757</td>
<td>6.912</td>
<td>0.692</td>
<td>0.490</td>
</tr>
<tr>
<td>2</td>
<td>USA 6.558</td>
<td>6.827</td>
<td>6.480</td>
<td>-58</td>
<td>-37</td>
<td>2.867</td>
<td>4.301</td>
<td>0.442</td>
<td>0.664</td>
</tr>
<tr>
<td>3</td>
<td>28 European Union Member States</td>
<td>4.572</td>
<td>4.759</td>
<td>-49</td>
<td>-41</td>
<td>2.427</td>
<td>2.808</td>
<td>0.607</td>
<td>0.702</td>
</tr>
<tr>
<td>4</td>
<td>India 2.772</td>
<td>2.886</td>
<td>5.450</td>
<td>98</td>
<td>84</td>
<td>5.714</td>
<td>5.310</td>
<td>1.048</td>
<td>0.974</td>
</tr>
<tr>
<td>5</td>
<td>Russia 2.230</td>
<td>2.321</td>
<td>2.650</td>
<td>-73</td>
<td>-48</td>
<td>0.627</td>
<td>1.207</td>
<td>0.237</td>
<td>0.456</td>
</tr>
<tr>
<td>6</td>
<td>Japan 1.251</td>
<td>1.302</td>
<td>1.210</td>
<td>-45</td>
<td>-50</td>
<td>0.716</td>
<td>0.651</td>
<td>0.592</td>
<td>0.538</td>
</tr>
<tr>
<td>7</td>
<td>Brazil 0.950</td>
<td>0.989</td>
<td>1.300</td>
<td>-45</td>
<td>-35</td>
<td>0.544</td>
<td>0.643</td>
<td>0.418</td>
<td>0.494</td>
</tr>
<tr>
<td>8</td>
<td>Indonesia 0.709</td>
<td>0.738</td>
<td>1.050</td>
<td>-32</td>
<td>-39</td>
<td>0.502</td>
<td>0.450</td>
<td>0.478</td>
<td>0.429</td>
</tr>
<tr>
<td>9</td>
<td>Canada 0.705</td>
<td>0.734</td>
<td>0.800</td>
<td>-70</td>
<td>-41</td>
<td>0.220</td>
<td>0.433</td>
<td>0.275</td>
<td>0.541</td>
</tr>
<tr>
<td>10</td>
<td>Mexico 0.704</td>
<td>0.733</td>
<td>0.850</td>
<td>13</td>
<td>-9</td>
<td>0.828</td>
<td>0.667</td>
<td>0.974</td>
<td>0.785</td>
</tr>
<tr>
<td>11</td>
<td>South Korea 0.659</td>
<td>0.686</td>
<td>0.730</td>
<td>-43</td>
<td>-54</td>
<td>0.391</td>
<td>0.316</td>
<td>0.536</td>
<td>0.432</td>
</tr>
<tr>
<td>12</td>
<td>Australia 0.560</td>
<td>0.583</td>
<td>0.670</td>
<td>-65</td>
<td>-30</td>
<td>0.204</td>
<td>0.408</td>
<td>0.305</td>
<td>0.609</td>
</tr>
<tr>
<td>13</td>
<td>Saudi Arabia 0.478</td>
<td>0.498</td>
<td>1.290</td>
<td>-38</td>
<td>-22</td>
<td>0.309</td>
<td>0.388</td>
<td>0.239</td>
<td>0.301</td>
</tr>
<tr>
<td>14</td>
<td>South Africa 0.459</td>
<td>0.478</td>
<td>0.940</td>
<td>-37</td>
<td>-33</td>
<td>0.301</td>
<td>0.320</td>
<td>0.320</td>
<td>0.341</td>
</tr>
<tr>
<td>15</td>
<td>Turkey 0.379</td>
<td>0.395</td>
<td>1.110</td>
<td>6</td>
<td>-5</td>
<td>0.418</td>
<td>0.375</td>
<td>0.377</td>
<td>0.338</td>
</tr>
<tr>
<td>16</td>
<td>Argentina 0.324</td>
<td>0.337</td>
<td>0.520</td>
<td>-24</td>
<td>-28</td>
<td>0.256</td>
<td>0.243</td>
<td>0.493</td>
<td>0.467</td>
</tr>
<tr>
<td>Total</td>
<td>33.075</td>
<td>34.430</td>
<td>43.150</td>
<td>-40</td>
<td>-41</td>
<td>26.082</td>
<td>25.431</td>
<td>0.604</td>
<td>0.589</td>
</tr>
</tbody>
</table>

Note: emissions after reduction = (2010 emissions) (1 + emissions reduction); and expected national demands = emissions after reduction/ (2030 emissions).
We find fair equilibria for both scenarios and examine how equilibrium demands for other nations are affected by the US defection (Figure 3). The exact impact of the US exit on other countries depends on the national disagreement value \(0 < \delta < 1\). For low disagreement values and under the ECPC50 allocation rule China has to decrease its fair equilibrium demand from \(d^* = 0.69\) (69% of 2030 emissions) to \(d^* = 0.50\) (50% of 2030 emissions) for a net 0.19 demand reduction as shown in Figure 3. In the same terms but under the CDC rule, China’s fair equilibrium demand decreases from 0.49 to 0.37.

If the disagreement value increases, the nations that are required to make the lowest demands (e.g., Russia, Canada, and Australia under the ECPC50 approach) refuse to cover the difference created by the United States’ withdrawal from the negotiations. In turn, the rest of the world nations must decrease their emissions demand to stay below the global emissions threshold.

For example, if \(\delta = 0.4\), China has to decrease its equilibrium demand even further: from \(d^* = 0.50\) to \(d^* = 0.44\). However, for higher disagreement values such as \(\delta > 0.45\), China is no longer willing to cut its emissions to compensate for the US withdrawal. Consequently, India, Mexico, and the 28 European Union Member States (especially, under the CDC allocation) must decrease their emissions further to avoid collapse of the global mitigation efforts. Higher disagreement values

---

**Figure 2.** The fair equilibrium demands under the equal cumulative per-capita for 1950 (ECPC50) fair allocation approach.

*Note:* fair demand is derived on the basis of ECPC50 (Meinshausen et al., 2015) and Climate Action Tracker 2030 emissions projections at the time of the 2015 Paris Agreement and current national policies (Cat Consortium, 2015).
generally favor the developed countries at the expense of the developing nations’ demands.

In the baseline scenario, the sum of national equilibrium demands inevitably exceeds the climate change threshold for $\delta = 0.61$, which leads to a complete failure of international climate negotiations. In this case, the only fair equilibrium in the game is for all countries to defect. Neither country is willing to cut its GHG emissions since the cost of climate change mitigation is greater than the expected cost of climate change adaptation.

Therefore, $\delta < 0.61$ represents an important theoretical condition for the prospects of successful climate change mitigation. If the US withdraws from the Paris Agreement, international negotiations collapse for $\delta = 0.54$. In practice, this value is likely to be substantially lower since we cannot expect the developing countries to single-handedly solve the problem of climate change mitigation.

### 6. Conclusion

Climate change mitigation is one the most important political problems of our time. Theoretically it lies at the intersection of a collective risk social dilemma (Milinski et al., 2008) and a bargaining problem (Nash, 1950a; Smead et al., 2014). The
The collective goal is to cut GHG emissions and avoid catastrophic global warming. This problem is made more complicated by the fact that national total emissions today and in the future may be orthogonal to national emissions per capita as well as historical, that is, cumulative, emissions. The nations with low emissions per capita and/or low historical emissions may demand a more favorable bargaining outcome on the basis of fairness.

The traditional approaches to bargaining do not take this new information into account since there is no fair allocation reference point in those models (Kalai and Smorodinsky, 1975; Nash, 1950a; Smead et al., 2014). In the present paper, we attempt to remedy this situation since such a reference point can, in fact, be found. Using the state-of-the-art findings from the effort-sharing literature (Meinshausen et al., 2015), we derive fair allocation reference points, which then can be used in a bargaining model. To include these considerations into a bargaining solution, we define ‘fair equilibrium’ as a set of strategies (here, emissions cuts) that minimize

Figure 3. Emissions demand reduction due to the US withdrawal from international climate negotiations.

Notes: the equal cumulative per-capita (ECPC) case represents the ECPC allocation rule (Bode, 2004; Meinshausen et al., 2015) while the common but differentiated convergence (CDC) scenario is the CDC convergence rule (Höhne et al., 2006; Meinshausen et al., 2015); demand reduction is calculated as the difference between: (a) the fair equilibrium demands if the US participates in international climate negotiations; and (b) the fair equilibrium demands under the US exit and its business-as-usual emissions demands (CAT Consortium, 2015); and the disagreement value $0 < \delta < 1$ is an exogenous parameter defined as the fraction of the demand the players get if they exceed the global emissions threshold (Smead et al., 2014).
the deviation from the fair allocation scenario subject to national disagreement values.

While the traditional model is based on a single set of exogenous parameters (the players’ disagreement values), the present model is based on both the disagreement values and the fair allocation reference point. However, this extension allows us to apply the model in, arguably, a more realistic setting than many existing models of climate change mitigation. The primary goal of the present paper is to explore the consequences of the US withdrawal from the Paris Agreement on climate change and answer the following two questions: Will the cooperative equilibrium still exist? Which nations will need to increase their contributions the most in order to fill the gap created by the US defection?

Although President Trump made his withdrawal announcement on June 1, 2017, there is still time to evaluate the consequences of such a decision. According to the terms of the Paris Agreement, the US President can give written notice of withdrawal as early as November 4, 2019, which would then take effect a year later, on November 4, 2020 (Bodansky, 2016).

In the present model, the impact of the US exit is (technically) analogous to lowering the climate change threshold for the rest of the world. All other things being equal, a lower threshold has a larger impact on larger countries. With that said, a lower threshold has a small impact on the countries which are close to the disagreement value—regardless of their size. Such countries are already expected to cut their emissions most aggressively—these are predominantly the developed countries. Therefore, all developed countries—big and small—are less affected by a lower threshold caused by the US withdrawal from the negotiations.

On the contrary, a lower threshold has a large impact on the countries which are not required to cut their GHG emissions most dramatically. According to both ECPC50 and CDC effort-sharing approaches, these are the developing countries. Since small developing countries have virtually no influence on the amount of global emissions, it is the large developing countries such as China, India, and Mexico that have to cut their emissions the most to compensate for the US defection. While such a scenario is technically possible, the history of climate negotiations (Bodansky, 2001) suggests that these developing countries may not be willing to accept an unfair distribution of the emissions cuts associated with the US withdrawal from international climate negotiations.

Acknowledgement
This paper was presented at the 7th Biennial Meeting of the Social Dilemmas Working Group, June 15–17, 2017, University of Massachusetts Amherst.

Funding
The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the National Science Foundation (award #0940822).
ORCID iD
Oleg Smirnov https://orcid.org/0000-0002-9633-910X

Supplemental material
Supplemental material for this article is available online.

References


