The Prisoner's Dilemma in Environmental Politics: One Model to Rule Them All?

https://www.e-ir.info/2018/06/14/the-prisoners-dilemma-in-environmental-politics-one-model-to-rule-them-all/

ANDREï-BOGDAN STERESCU, JUN 14 2018

In this essay I will look at how, and if so, in what way, does the Prisoner's Dilemma (PD) explain the character of International Environmental Politics (IEP) and International Environmental Agreements (IEA). I will start by looking at how the PD pertains to the international system and environmental politics through the logic of collective action and Social Dilemmas like the Free Rider problem and the Tragedy of the Commons. I will then show how an abstract model of the PD, repeated PDs, and the Tit-for-Tat strategy can be applied to different theoretical scenarios in IEP, and look at the strengths and weaknesses of the model. I will finish by contrasting the PD with some alternative game theoretic and international negotiation and resolution models for analyzing IEP and IEA like the Stag Hunt/ Cooperation Game and Putnam's Two-Level Game Theory which could replace or complement the PD in analyzing the dynamics of IEP and IEAs.

The International System and the Logic of Collective Action

In an article from 2006, Stephen M. Gardiner indentifies three characteristics of the climate change problem, and also other environmental problems, which constitutes "The Global Storm" (Gardiner, 2006): Dispersions of Causes and Effects, Fragmentation of Agency, and Institutional Inadequacy.

Gardiner rightly observes that in the case of the 2nd characteristic, the structure of agency is understood to be arising out of the shape of the international system, as constituted by states, but also by institutions, non-state actors, etc. Gardiner also says that the international situation is generally understood through the prism of a "familiar theoretical model", Game Theory. The author employs the two most common iterations of this model when it comes to International Environmental Politics: Prisoner's Dilemma and the Tragedy of the Commons.

The Tragedy of the Commons as developed by Garrett Hardin (Hardin, 1968) is usually taken to be an example of the Prisoner's Dilemma, because it is a problem of collective action, but the Tragedy of the Commons is illustrative of failed cooperation scenarios in general and it lends itself to various Game Theoretic models. Although it has been observed in different contexts across history, the problem of collective action was analytically laid down by economist Mancur Olson in his 1965 book, the Logic of Collective Action (Olson, 1971). The collective action problem is a situation in which two or more parties desire and would benefit from a common good, for example a cleaner atmosphere, but individual interest acts against the common interest. This means that actors won't take individual action towards achieving the common good.

Social Dilemmas in International Environmental Politics

The logic of collective action is illustrated by two very similar social dilemmas. A social dilemma is a situation in which a non-cooperative outcome is sub-optimal. (Wood, 2011) The first is the Free-Rider problem and the second is Hardin's Tragedy of the Commons. The former illustrates a situation in which actors benefit from a common or a public good, but do not contribute towards its maintenance or its production. The latter refers to a situation in which individual consumption of a common or a public good leads to the degradation of said good, interfering with the ability of others to enjoy it.

As said above, usually a problem of collective action, and implicitly the strategic structure of a social dilemma, can be modeled as an n-player PD. (Hardin, 1971) If n=2, figure 1 shows the pay-off matrix of a non-cooperative game in which the outcomes are ranked worst (1) to best (4). Essentially, the PD refers to a situation in which two players have to choose between two options and the outcome of each depends on the simultaneous decision of the other.

Consider a climate negotiation game where A and B are two countries that face the choice to either mitigate pollution, reducing GHG emissions, or not reducing GHG emissions, leading to more pollution. The choice to mitigate by one of the countries would benefit both, but a policy of climate mitigation assumes certain costs in terms of economic strength and geopolitical competition. Both players have a Dominant Strategy, one that is better irrespective of what the other player does.

		Country A: Reduce Emissions	Country A: Don't Reduce Emissions
Country B	Reduce Emissions	3, 3	1, 4
	Don't Reduce Emissions	4, 1	2, 2

Figure 1: **One-Shot Prisoner's Dilemma Pay-Off Matrix**

For both A and B the Dominant Strategy is to not reduce emissions, and usually in non-cooperative games it is assumed that the Dominant Strategy will be chosen. Why is this so? If Country A chooses to reduce emissions while country B continues with business as usual, then Country A will be in a position of structural disadvantage while country B will benefit from country A's choice without incurring any immediate costs. If both countries choose to not reduce emissions, then they both experience the penalties of continuing with business as usual, but suffer no immediate decline in terms of economic strength and geopolitical competition (meaning that near-term costs deter countries from achieving distant benefits).

In this case the Dominant Strategy is also a Nash Equilibrium, since neither of the two players would benefit from deviating unilaterally from the outcome. In Game Theoretic terms, the choice to not reduce emissions by both countries is a Pareto-Inferior outcome. To relate this back to the common interests vs. personal interest problem, both countries are better off collectively if they mitigate, but they are individually better off if they pollute.

The Prisoner's Dilemma in International Environmental Politics

The character of International Environmental Politics has often been said to resemble a Prisoner's Dilemma, same with Climate Negotiations. (Gardiner, 2006; van der Gaast, 2017; Liebreich, 2007; Sooros, 1994; Endres, 2004; Barrett & Dannenberg, 2012) Analysts, scientist, and economists have applied the model to different scenarios like the now defunct Kyoto Protocol in International Politics, which is taken to be an example of a deal that incentivized free-riding behavior (Liebreich, 2007; The Economist, 2007; Weiler, 2010; Barrett, 2014). The impasse over the Kyoto Protocol seemed to resemble a static PD at the moment, just like in *Figure 1*. However, Michael Leibreich suggests seeing not only the Kyoto Protocol, but IEAs in general, as a repeated PDs. (Liebreich, 2007) Building on Robert Axelrod's study of how cooperation can emerge and persist (Axelrod, 2006), Liebreich suggests that the outcome can change since, in the real world, countries can re-assess whether to reduce emissions or not in subsequent negotiations and agreements. The idea is that the best strategy in a repeated PD would be a "tit-for-tat"

strategy:

1. Players start by cooperating
2. If a player defects, then he should be punished in subsequent games until cooperation is reinstated
3. Cooperation is reinstated

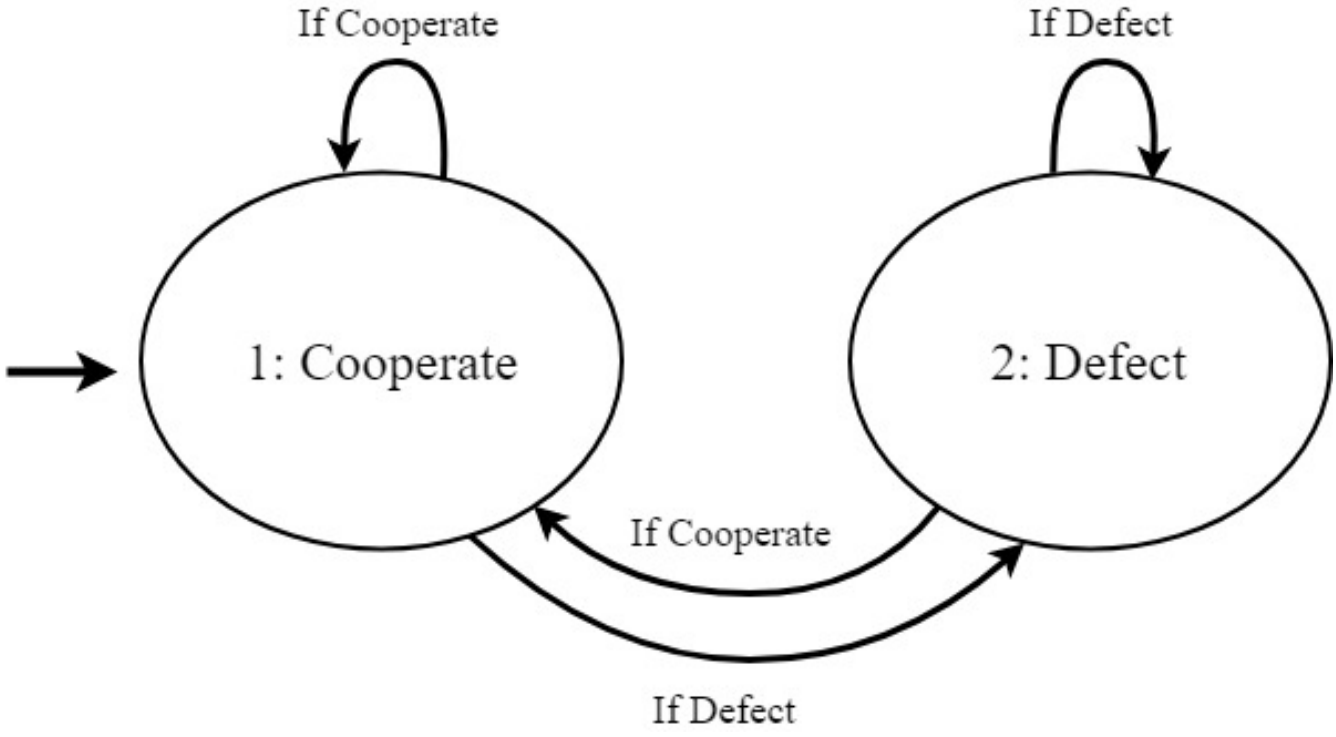

Figure 2: **Simplified structure of a Tit-for-Tat Strategy in a Repeated PD**

In the real world cooperation is more likely when there is room for communication, and the likelihood of cooperation increases in subsequent games. A Tit-for-Tat strategy also presupposes that players take into account each other's previous decisions. Cooperative behavior is also more likely in repeated games if players have long-term perspectives and also a mechanism for punishing defection. (Osborne & Rubinstein, 1994; Finus, 2001; Wood, 2011; Endres, 2004)

Actors in the international system take decisions over time and not simultaneously, meaning that a repeated PD is a much better framework for analyzing IEP and IEA than a single-shot PD. Communication and constant interaction increase the likelihood of cooperation. However, this does not mean the model is not limited. One of its biggest drawbacks when applied to IEP and greenhouse gas abatement is that damages from GHG emissions come from cumulative emissions rather than emissions in a given period, meaning that actors cannot defect in subsequent games without causing lasting damage. (Wood, 2011) The pay-off of defection will eventually diminish to zero. Another drawback, as shown by Scott Barrett, is that when *n* becomes large, cooperation becomes more difficult to sustain. An IEA will support a large number of signatories when gains to cooperation are small (because incentives to free-ride are reduced). When gains are large, a self-enforcing IEA – meaning that no actor is forced to sign it and can withdraw at any time – can only sustain a small *n*. This means that in the absence of a binding commitment to reduce emissions, actors will eventually defect. (Barrett, 1994)

More so, the logic of a Tit-for-Tat strategy in iterated PDs rests on the fact that actors can monitor defections and can adequately respond to them. It's very difficult to keep track of free-riders and defectors in order to apply punishment without causing lasting harms to multiple actors. Also, without taking in account inequalities in power between

countries, it's very improbable that actors could apply punishment unilaterally or even multilaterally in the face of greater powers like the United States or China were they to defect on an IEA.

However, even though the PD model theoretically explains why actors tend to continue to pollute in the absence on an overarching regulatory body, it is an oversimplified explanation of IEP and game theory in International Politics. International negotiations on climate change, exploitation of natural resources, environmental issues, and common goods involve not two, but hundreds of state and non-state actors with their own preferences, interests and values. More so, given the nature of environmental problems, and especially Climate Change, the pay-offs of cooperation and defection cannot be known for sure in a real world scenario. It might just be so that cooperation would be more likely once we arrive at better approximations of those pay-offs.

In an article from Ecological Economics, DeCanio and Fremstad (2011) show that the PD might not be the best description of the character of climate negotiations by exploring different simple game theoretic models in IEP. Whilst an insightful model for why negotiation and cooperation are vital for the best outcomes, the PD, like any other model, is unrealistic in the sense that it fails to cover many aspects of the real world. (Endres, 2004) Even though negotiations and interactions between self-interested actors have been primarily modeled as PDs (Smead, et al., 2014; Barrett & Dannenberg, 2012), the fact that other models can explain the logic and character of IEP and IEAs means that the PD's explanatory power and ability to offer solutions is not so absolute when dealing with interactions between players at the international level in regards to environmental issues.

In its defense though, all models are simplified representations of reality and are thus limited. As George Box said: "all models are wrong; the practical question is how wrong do they have to be to not be useful?" (Box & Draper, 1987)

Alternative Game Theoretic Models in International Environmental Politics

Gardiner himself recognizes the limitations and incompleteness of the PD model, mentioning other models like the Battle of the Sexes, also called the Stag Hunt or the Cooperation Game. Although he recognizes alternative models in explaining the character of IEA, Gardiner is skeptical in regards to them, stating that "none of the main claims of the broader Battle of the Sexes model…seems likely to be true of climate change". (Gardiner, 2006; Smead & Sandler, 2013) While Gardiner might be skeptical about the Stag Hunt/Battle of the Sexes/Cooperation Game model (I will refer to it as the Stag Hunt), I believe that it is worth taking into consideration. Just like Mead and Sandler say, many of the potential solutions to the PD involve changing the incentives and pay-off structure of the game as to promote cooperation, which would turn it into a Stag Hunt in a way that is different from an Iterated PD.

Taking the example of the Tragedy of the Commons to the real world, our ecosystem and atmosphere are constantly degrading as we exploit and pollute them. As said above, the pay-off of defection will eventually diminish to almost zero. It is in this situation especially that the Stag Hunt model becomes more appealing. As said above, uncertainty about the pay-offs of both cooperation and defection might be the causes for the lack of cooperation so far.

The stag hunt is a story by Jean Jacques Rousseau that became a game. (Rousseau, 1985) In this game, two hunters are engaged in a stag hunt, and they can either stick to their post (cooperation) or hunt hare (defection). Hunting stag offers a higher pay-off if the other player cooperates, but nothing at all if he doesn't. Hunting hare provides a moderate but certain pay-off irrespective of the decision of the other player. (Smead & Sandler, 2013) Unlike the PD, the Stag Hunt has two Nash Equilibria, one that is risk dominant (Don't Reduce Emissions in *Figure 3*), and one that is pay-off dominant (Reduce Emissions).

		Country A: Reduce Emissions	Country A: Don't Reduce Emissions
Country B	Reduce Emissions	4, 4	1, 3
Country B	Don't Reduce Emissions	3, 1	2, 2

Figure 3: **One-Shot Stag Hunt Pay-Off Matrix**

In *Figure 3* both countries are highly averse to not reducing emissions, even if the adversary chooses to do so. (DeCanio & Fremstad, 2011) This actually reduces incentives to defect, given that an agreement to reduce emissions can be reached. Even if we assume that both countries would cut emissions unilaterally rather than cooperating, the matrix could be adapted to say that given the increasing costs of mediating climate change, choosing to cut emissions unilaterally would yield less benefits than doing so by working together (risk dominant strategy). The Stag Hunt is fundamentally different from the PD in the sense that the best outcome in the PD is to defect if the other player cooperates. In a Stag Hunt the best outcome is the pay-off dominant outcome.

When answering the question of whether the character of IEP and Climate Change is best described by the PD or the Stag Hunt, it depends on how severe the risk posed by Climate Change and other forms over-exploitation of resources is. DeCanio and Fremstad (2011, p. 182) say that the comparison between the PD and the Stag Hunt show that "the overriding barrier to achieving an international agreement to protect the climate may be a failure of the leading governments to grasp the seriousness of the climate risk." The two authors continue to present other simplified 2×2 models like the Chicken Game in their paper.

Scott Barrett and Astrid Dannenberg present a very interesting parallel between the Stag Hunt and the PD in a proceeding for the National Academy of Sciences of the United States. What the authors are saying is that ever since the Framework Convention on Climate Change was adopted in 1992, negotiations over emission limits have been intertwined with efforts to identify a critical threshold. When the threshold was indentified in the 2009 Copenhagen Accords, the 2°C threshold, its identification did not improve negotiations by a lot.

Basically, the idea of the authors is that collective action fails precisely because of uncertainty about the threshold. This uncertainty derives from the fact that scientific literature presents many views about a temperature threshold for "dangerous" climate change, and the fact that countries can only control emissions directly while the effect of emissions on temperature is uncertain. (Barrett & Dannenberg, 2012) Under uncertainty, countries that would otherwise play a Coordination Game will divert back to the Prisoner's Dilemma.

Two-Level Games in International Environmental Politics

One last model that I want to talk about is more of a method for analyzing specific negotiation scenarios rather than a model for explaining the character of IEP and IEA. Building on Robert Putnam's Two-Level Game theory (Putnam,

1988), we can imagine a Bargaining model where actors consider only those agreements that can be ratified by all parties, including at the national level. Given the fact that multiple pressure groups with different interests exist within a country, they can constrain action at the International level. (Kroll & Shogren, 2008) The way it works is that at the national level, domestic groups pursue their own agenda and can pressure the government into adopting favorable policies. At the international level, governments seek to maximize their ability to satisfy domestic pressures, while also minimizing the adverse effects of foreign negotiations. At Level I, negotiations between international actors lead to a tentative agreement, and at Level II the dynamics between domestic groups and the government can lead to either the formal or informal ratification of the tentative agreement at Level I, or its abandonment. Putnam defines "win-sets" as the set of all possible Level I agreements that can be ratified at Level II. Larger win-sets make agreements at Level I more likely to be ratified at Level II.

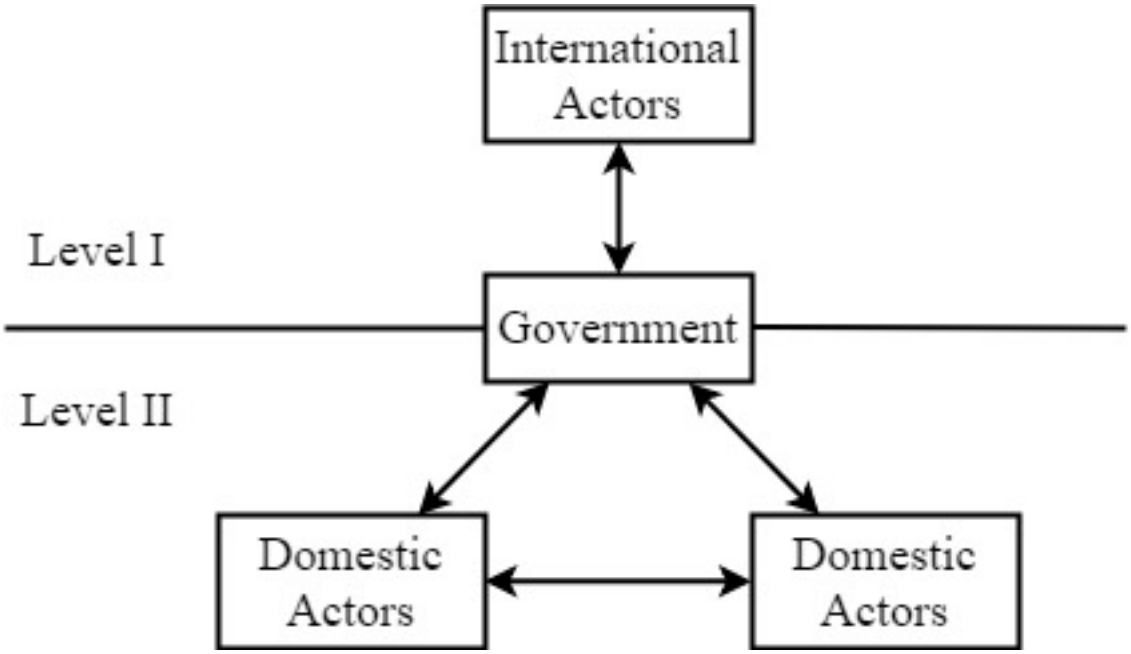

***Figure 4*: Structure of a Two-Level Game**

The Kyoto Protocol again can be analyzed from the perspective of a Two-Level Game. While at the International level Bill Clinton and Al Gore negotiated an agreement that would make the administration seem climate-friendly, Gore's commitment to a 7% reduction in US emissions could not pass a Senate that had just approved Byrd-Hagel Resolution, which went in the opposite direction. (Amundsen & Lie, 2010) In 2001, during George W. Bush's presidency, the US officialy withdrew from the Kyoto Protocol. Michael Lisowski (2010) argues that President Bush repudiated the Protocol by appealing to the Byrd-Hagel Resultion and securitizing California's energy crisis in order to pursue his own domestic energy policy, making it impossible to engage in further negotiations at the international level. Looking at the Kyoto Protocol from a Two-Level Game perspective allows us to see that further negotiations, even if problems related to free-riding would have been accounted for, would have not probably led to an agreement that could have been ratified by the US Senate because such an agreement was not included in any of the US win-sets.

Robert Keohane and Michael Oppenheimer also analyze the situation of OECD, BRICS and poorer states from the perspective of a Two-Level Game. For example, the principle focus of BRICS governments is economic growth, but since they are affected by Climate Change (or just because they see industrial opportunities in global action) they do have reasons to act on the issue of Climate Change. The Two-Level Game implies the fact that they seek to do enough to induce action by others while also avoiding sanctions against themselves, but not so much that it would burden their economic growth or anything that reduces public support for the ruling group (e.g. China). (Keohane & Oppenheimer, 2016)

Conclusion

What the Prisoner's Dilemma shows us with little doubt left is that IEP tend to have winners and losers, but it also points out that the best collective decision would be for all actors in IEP to cooperate and commit to reducing emissions, or to make an agreement as to avoid the overexploitation of common resource pools and overpollution. The PD offers an insightful explanation for why is it so hard for actors in IEP to give in, and it also provides a frame in which to solve the problem that the dilemma itself produces. Whilst a simplified PD model with a high level of abstraction would not incorporate the complexity of IEP, the logic of the game seems to remain strong in regards to actors' incentives to defect rather than cooperate when dealing with situations like the Tragedy of the Commons or the Free-Rider problem in IEP and IEA and it provides a coherent explanation for the current impasse in IEP.

But, whilst insightful in showing the current diplomatic impasse in IEP, "the important decision-theoretic features of the climate problem are not exhausted by the Prisoner's Dilemma". (Smead & Sandler, 2013, p. 19) If there is something that I have shown is this essay is that the PD is neither the sole, nor necessarily the best explanatory or predictive model when it comes to IEP and IEA. Changes in the way we conceptualize the problem, as well as the pay-off matrix and our current disposition in the real world can lead to different ways to understand the character of IEP, and there is at least enough evidence to support alternative models as there is supporting the PD.

Bibliography

Amundsen, B., & Lie, E. (2010). *Why the Kyoto agreement failed*. Retrieved January 6, 2018, from https://www.forskningsradet.no/en/Newsarticle/Why_the_Kyoto_agreement_failed/1253963392536?lang=en

Axelrod, R. (2006). *The Evolution of Cooperation* (Revised ed.). Basic Books.

Barrett, S. (1994). Self-Enforcing International Environmental Agreements. *Oxford Economic Papers , 46*, 878-894.

Barrett, S. (2014). Why Have Climate Negotiations Proved So Disappointing? . Vatican City: The Pontifical Academy of Social Sciences.

Barrett, S., & Dannenberg, A. (2012). Climate negotiations under scientific uncertainty. *Proceeding of the National Academy of Sciences of the United States , 109* (43), 17372–17376.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (1st ed.). Wiley.

DeCanio, S. J., & Fremstad, A. (2011). Game theory and climate diplomacy. *Ecological Economics , 85* (1), 177-187.

Endres, A. (2004). Game theory and global environmental policy. *Poiesis & Praxis , 3* (1-2), 123–139.

Finus, M. (2001). *Game Theory and International Environmental Cooperation* (1st ed.). Edward Elgar Publishing .

Gardiner, S. M. (2006). A Perfect Moral Storm: Climate Change, Intergenerational Ethics and the Problem of Moral Corruption. *Environmental Values , 15* (3), 397-413.

Hardin, G. (1968). The Tragedy of the Commons. *Science , 162* (3859), 1243-1248.

Hardin, R. (1971). Collective action as an agreeable n-prisoners' dilemma. *Systems Research and Behavioral Science , 16* (5), 472–481.

Keohane, R. O., & Oppenheimer, M. (2016). Paris: Beyond the Climate Dead End through Pledge and Review? *Politics and Governance , 4* (3), 142-151.

Kroll, S., & Shogren, J. F. (2008). Domestic politics and climate change: international public goods in two-level games. *Cambridge Review of International Affairs , 21* (4), 563-583.

Liebreich, M. (2007). *How to Save the Planet: be Nice, Retaliatory, Forgiving and Clear.* New Energy Finance. Retrieved January 4, 2018, from https://about.bnef.com/blog/how-to-save-the-planet-be-nice-retaliatory-forgiving-and-clear/

Lisowski, M. (2010). Playing the Two-level Game: Us President Bush's Decision to Repudiate the Kyoto Protocol. *Environmental Politics* , 101-119.

Olson, M. J. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups* (2nd Revised ed.). Harvard University Press.

Osborne, M. J., & Rubinstein, A. (1994). *A Course in Game Theory* (1st ed.). MIT Press.

The Economist. (2007, September 27). Playing Games with the Planet. Retrieved January 4, 2018, from https://www.economist.com/node/9867020

Putnam, R. (1988). Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization , 42* (3), 427-460.

Rousseau, J. J. (1985). *A Discourse on Inequality* (Reprint ed.). Penguin Classics.

Smead, R., & Sandler, R. (2013). Game Theory and the Ethics of Global Climate Change .*Philosophy and Public Issues , 3* (1), 13-23.

Smead, R., Sandler, R. L., Forber, P., & Basl, J. (2014). A bargaining game analysis of international climate negotiations. *Nature , Climate Change* (4), 442–445.

Sooros, M. S. (1994). Global Change, Environmental Security, and the Prisoner's Dilemma. *Journal of Peace Research , 31* (3), 317-332.

van der Gaast, W. (2017). *International Climate Negotiation Factors: Design, Process, Tactics* (1st ed.). Springer International Publishing.

Weiler, F. (2010). *Global Climate Change and Leadership: The Role of Major Players in Finding Solutions to Common Problems*. Retrieved January 4, 2018, from http://www.saisjournal.org/posts/global-climate-change-and-leadership

Wood, P. J. (2011). Climate Change and Game Theory: A Mathematical Survey.*CCEP Working Paper No. 2.10 , Available at SSRN: https://ssrn.com/abstract=1883944 or http://dx.doi.org/10.2139/ssrn.1883944*.

Written by: Andreï-Bogdan Sterescu
Written at: University of Birmingham
Written for: Dr Emma Foster
Date written: January 2018