Catharina Lee

Gaming our way out of the climate breakdown – How can game theory models support the negotiations of more ambitious climate agreements?

Faculty of Law

Victoria University of Wellington

2022

Contents (References - table of contents)

I Introduction

"All models are simplified versions of reality, full of limitations".[1] Game theory models applied in the context of the international climate negotiation game are no different. Every model of the climate negotiation game is certain to lack factors integral to its accurate characterisation. This is because at its essence, "all models are wrong, but some are useful."[2] The practical question then is, how wrong must a game theory model be to not be useful? This question is considered in the context of the international climate negotiation game throughout this paper.

After introducing game theory in Part 2, this paper considers game theory within the international climate negotiation context. Part 3 explores the idea of climate change as a grave example of the common-pool resource problem known as the "tragedy of the commons". The general utility and limitations to game theory in this context are also discussed. In Part 4, the key assumptions, and issues that an international climate negotiation game must consider are outlined. Part 5 considers how four traditional game theory models can be applied to the international climate negotiation game. Part 6 discusses factors that could aid the design of an international climate negotiation game, and mechanisms that will facilitate a cooperative outcome.

This paper concludes that although game theory models can never be complete nor fully accurate due to the underlying complexities and uncertainties in the international climate negotiation game,[3] a useful model need only be "not wrong enough" to be of use. The final part of this paper considers the persuasive argument which threatens to disgruntle many collective action scholars, that distributive politics better characterises the climate

---

[1] Kaveh Madani "Modeling international climate change negotiations more responsibly: Can highly simplified game theory models provide reliable policy insights?" (2013) 90 Ecological Economics 68 at 68.
[2] George E.P Box and Norman R. Draper "Empirical model-building and response surfaces" (1987) John Wiley & Sons in Madani, above n 1, at 68.
[3] Shi-Ling Hsu "A Game Theoretic Model of International Climate Change negotiations" (2010) 19 NYU Environmental Law Journal 14 at 31.

negotiation game at hand.[4] Collective action has been posited as providing the wrong solution to a misunderstood problem for too long.[5] Hence, we consider that Hale's 'Catalytic Cooperation" model, which combines the best of the collective action and distributive politics approaches, provides the most promising way forward.[6]

II Game Theory

A What is Game Theory?

Game theory is a part of the body of mathematics which models and analyses how economic agents interact, how outcomes are produced with respect to the various preferences (utilities) of those agents.[7] It is often used as a theoretical framework to conceive social situations among competing players.[8] The focus is on the game, the set of circumstances that form a model of the interactive situation, and the players, the strategic decision-makers. Game theory models diverge from each other by identifying different players, preferences, available strategies, payoffs, and potential outcomes. Varying requirements and assumptions may also be present, depending on the model. The choice of action for each participant in the game theory model will duly affect the outcome.

There are certain assumptions present in game theory generally, and limitations to the applicability of specific game theory models to real life.[9] In terms of general game theory assumptions, first, all players within a game are assumed to be utility-maximising rational

[4] Michaël Aklin and Matto Mildenberger "Prisoners of the Wrong Dilemma: Why Distributive Conflict, Not Collective Action, Characterizes the Politics of Climate Change" (2020) 20(4) Global Environmental Politics 73.
[5] Aklin and Mildenberger, above n 4, at 79; and David G. Victor *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet* (Cambridge University Press, Cambridge, UK, 2011).
[6] Thomas Hale "Catalytic Cooperation" (2020) 20(4) Global Environmental Politics 73.
[7] "Game Theory" (8 March 2019) Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/entries/game-theory/index.html>.
[8] Adam Hayes "Game Theory: Definition, Role in Economics, and Examples" (2 February 2022) Investopedia <https://www.investopedia.com/terms/g/gametheory.asp>.
[9] Hayes, above n 8.

actors who will strive to maximise their payoffs in the game.[10] This paper will adopt the understanding of the economically rational player discussed in the Stanford Encyclopedia of Philosophy: The economically rationally player is one who can assess outcomes and rank-order them with regard to their utility, calculate paths to the outcomes by associating actions with their outcomes, and select actions from sets of alternatives.[11] Second, in the simplest games, players are assumed to have full information about the game, its rules, and the consequences.[12] This is known as perfect information – the agent knows everything that has occurred in the game up until the point where they must take an action.[13] Third, players cannot communicate or interact with each other.[14] Consequently, possible outcomes are known in advance and cannot be changed.

The above assumptions provide parameters and a framework to understand real life situations through the lens of game theory models. However, they also serve to limit the applicability of game theory models in different situations, where the assumptions do not align with reality.[15] Awareness of the assumptions within game theory models and their implications when applied to real world situations are essential when considering how game theory can support the negotiation of more ambitious climate agreements. The assumptions upon which a model rests should capture the essence of the situation, excluding irrelevant details.[16]

Before discussing game theory models in-depth, a basic understanding of the key terms commonly used in game theory is helpful. The strategic game is a set of circumstances that has a result dependent on the actions of two or more decision-makers (players), who are

---

[10] See the discussion about the "theory of rational choice" in Martin J. Osborne *An Introduction to Game Theory* (Oxford University Press, 2003) at 2; and Steven Tadelis *Game Theory: An introduction* (Princeton University Press, United Kingdom, 2013) at 11.
[11] Stanford Encyclopedia of Philosophy, above n 7.
[12] See the general discussion of games with perfect or complete information by Osborne and Tadelis, above n 10.
[13] Stanford Encyclopedia of Philosophy, above n 7.
[14] Hayes, above n 8.
[15] Osborne, above n 10, at 1.
[16] Osborne, above n 10, at 1.

the agents involved in playing the game.[17] Each player will have a set of actions and preferences over the set of action profiles in the game.[18] A strategy is the predetermined plan of action or 'programme of play' that a player undertakes, in response to the possible strategies other players may use.[19] The payoff is what the player will receive from a particular outcome, it represents the decision-maker's preferences.[20] This can be in any quantifiable form, whether that is monetary value or utility generally.[21]

Regarding the outcomes (the possible consequences that can result from any of the actions),[22] equilibrium is the point in the game where all players have made their decisions and an outcome is reached.[23] Nash equilibrium is the point where no player can increase their payoff or do better by changing their decision unilaterally, given that every other player adheres to their chosen action.[24] A subgame perfect equilibrium is the strategy combination constituting a Nash equilibrium in every subgame of the overall game.[25]

B Purpose of Game Theory

Game theory is not only useful for economic purposes but can also be abstracted to real-life situations to critique human rationality and understand human responses to certain incentives.[26] Game theory has been described as the 'science of strategy'.[27] Using game theory, real world situations can be laid out and their outcomes can be predicted. At its essence, game theory can be used to provide an explanation of the strategic reasoning process that humans undergo. The study of game theory models and the process of understanding how individual actors adopt strategies particular to different situations with

---

[17] Osborne, above n 10, at 11; and Hayes, above n 8.
[18] Osborne, above n 10, at 11.
[19] Stanford Encyclopedia of Philosophy, above n 7.
[20] Osborne, above n 10, at 4-5; and Tadelis, above n 10, at 7.
[21] Hayes, above n 8.
[22] Tadelis, above n 10, at 4.
[23] Hayes, above n 8.
[24] Osborne, above n 10, at 20.
[25] Peter John Wood "Climate change and game theory" (2011) 1219 Ann. N.Y. Acad.Sci 153-170 at 157.
[26] David Mond "Game theory and climate change" in *The Impacts of Climate Change* (Elsevier, 2021) 437 at 437.
[27] Hayes, above n 8.

varying levels of utility present, has a significant psychological element.[28] This is, however, as aforementioned, constrained by assumptions such as that certain actions are the ‘rational’ ones, and that the agent will always follow the ‘rational’ option. These assumptions may not always be consistent with experimental psychology.[29]

III Game Theory in Climate Change

This section discusses the utility and purpose of game theory in the climate change context. First, it introduces the concept of climate change as an extreme example of the “tragedy of the commons”. It explains how game theory models can be used to analyse and provide important insights into the present situation regarding climate change and proposes ways that game theory can be used to incentivise climate mitigation. Next, it points out limitations to the utility of game theory in the climate context. Key constraints include the general inherent assumptions in game theory models, the risks of over-simplifying a complex problem, the high levels of uncertainty associated with climate change, and the urgent need for more information.

A Climate change as the “Tragedy of the Commons”

Climate change has traditionally been understood using the game theory representation of the “tragedy of the commons”, a metaphor for the tension between parties pursuing individual selfish behaviour or the collective interest.[30] The common pool resource problem, a microeconomic concept, was first introduced by Garret Hardin in his 1968 article as the ‘tragedy of the commons’.[31] Hardin observed that users of a shared resource have an incentive to overconsume. Where villagers graze their animals on common land, each rational individual villager has an incentive to add more of their animals onto the land, until the land becomes overgrazed and is ruined. This same temptation is present when we consider climate change. We can conceptualise the climate, the average level of temperature in a region, as a shared resource that is necessary for healthy living, cultivating

[28] Colin F. Camerer “Progress in Behavioral Game Theory” (1997) 11 Journal of Economic Perspectives 167.
[29] Wood, above n 25, at 154.
[30] Karl-Henrik Robert and Goran Broman "Prisoners' dilemma misleads business and policy making" (2017) 140 Journal of Cleaner Production 10-16.
[31] Garret Hardin “The Tragedy of the Commons” (1968) 162 Science 1243.

land, transportation etc, but which can also be ruined by excessive carbon-dioxide emission.[32]

Climate change provides an extreme example of the tragedy of the commons,[33] where individual actors are always tempted to increase their use of shared resources. The benefit of activities such as burning fossil fuels, cutting down forests and farming livestock go principally to the perpetrators, whilst the effects of it are suffered by all. Conversely, any effort to reduce emissions is at significant cost to the author, while the benefit is shared by all. This leaves open the possibility and indeed the incentive of individuals relying on other actors to, for example, reduce emissions, without needing to contribute themselves. This is commonly referred to as 'free-riding'. In the optimum situation, those sharing a common pool resource would agree on how to sustain and finance the resource, ensuring that it is maintained for the good of all.[34] However, through the lens of the rational individual model of microeconomics (as recognised by Olson), an actor who cannot be excluded from using a resource will not be motivated to contribute to the financing or sustaining of it, but will behave as a free-rider.[35] When individuals, corporations, and nations seek to increase their use of the shared resource of our climate without limit, in a world with limits, the damage caused can be felt throughout the world.

B How useful is game theory in climate change?

Game theory models play an essential role in helping us to analyse and understand the strategic behaviour of actors in international climate negotiations.[36] This includes understanding actors' incentives to freeride, the potential barriers to cooperation, and how a cooperative outcome should be approached and facilitated.

---

[32] Gábor Kutasi "Climate Change in Game Theory Context" (2012) 13(1) Interdisciplinary Environmental Review 42 at 47.
[33] Mond, above n 26, at 445.
[34] Kutasi, above n 32, at 47.
[35] Mancur Olson *The logic of collective action: public goods and the theory of groups* (Harvard University Press, 1965) in Kutasi, above n 32, at 47.
[36] Wood, above n 25, at 153; and Jobst Heitzig, Kai Lessmann and Yong Zou "Self-enforcing strategies to deter free-riding in the climate change mitigation game and other repeated public good games" (2011) 108 Proceedings of the National Academy of Sciences at 15739.

Game theory models can provide important insights into understanding states' participation and compliance with international climate agreements. For example, game theory can be used to analyse the failure of the 2009 Copenhagen climate negotiations and the resultant "weak" political agreement known as the "Copenhagen Accord".[37] The Accord was described as a broad but shallow agreement, that demanded little of each country but achieved universal participation.[38] It allowed countries to individually set their own greenhouse gas emission reduction goals, but it failed to set basic targets for reducing the global annual emissions of greenhouse gases and failed to secure commitments from countries to meet the targets collectively.[39] The lack of political will of major emitters, showcased by countries like the United States wanting voluntary "pledge and review" type of agreements with minimum enforcement mechanisms, resulted in weak environmental outcomes overall.[40] Furthermore, the lack of aggressive climate action taken by large emitters like the United States resulted in other nations being less inclined to increase their pledges to reduce emissions, for fear that this would result in jobs being exported to the more laxly regulated United States.[41]

Using game theory, the tension between environmental outcomes and participation by various parties in international agreements can be explored. In the previous example, there was wide participation in the Copenhagen Accord, but the US' decision to defect and free-ride off other nations' efforts can explain why other players defected eventually too, leading to the overall failure of the Accord. This type of agreement and the resultant outcomes are typical of the Prisoner's Dilemma game theory model, which conceptualises a game where players have little to lose by participating in the game but are incentivised to

---

[37] "*Summary of the Copenhagen Climate Change Conference: 7-19 December 2009* (International Institute for Sustainable Development: Earth Negotiations Bulletin, December 2009).

[38] Brian Spak "The Success of the Copenhagen Accord and The Failure of the Copenhagen Conference" (Substantial Research Paper, 2010) at 40.

[39] Adam Vaughan and David Adam "Copenhagen climate deal: Spectacular failure – or a few important steps?" (22 December 2009) The Guardian <https://www.theguardian.com/environment/2009/dec/22/copenhagen-climate-deal-expert-view>.

[40] Vaughan and Adam, above n 39.

[41] Spak, above n 38, at 40.

free-ride off others' contributions. The lack of agreement and strong incentive to free-ride resulted in parties defecting and poor overall climate outcomes being reached. This example reflects how game theory can help us to understand the strategies and incentives facing various parties, to explain the outcome in climate negotiations that was reached, and to provide opportunities to prevent similar situations occurring again.

Game theory models can also be used to investigate the political process of decision-making within and by different states.[42] Stakeholder's hesitations, rooted firmly in their uncertainty as to their counterpart's strategies, can be better comprehended and combatted.[43] Game theory models help us to identify the incentives which might prevent individuals from changing their behaviour.[44] The models can forecast the strategy of various stakeholders,[45] offering a glimpse into how situations might end, which can have profound long-term effects.[46] Understanding the strategies of different parties at play and utilising this information to inform and guide climate negotiations is likely to have significant policy implications domestically,[47] as well as impacting the design of international climate agreements overall.

Not only does game theory provide an opportunity to analyse already concluded climate negotiations, but it can play a significant role in incentivising action. Implementation theory is an area of game theory that treats the game as something to be *designed*, rather than as a given.[48] The desired outcome is treated as a given, and a reverse engineering approach is used to design a process to lead to that outcome.[49] The strategic properties of different games are analysed to assess what might induce individuals to always choose

---

[42] Wood, above n 25, at 154.
[43] Kutasi, above n 32, at 42.
[44] Wood, above n 25, at 70.
[45] Kutasi, above n 32, at 42.
[46] António Bento Caleiro, Miguel Rocha de Sousa, and Ingo Andrade de Oliveira "Global Development and Climate Change: A Game Theory Approach" in Tiago Sequeira and Liliana Reis (eds) *Climate Change and Global Development: Contributions to Economics* (Springer, Cham, 2019).
[47] Wood, above n 25, at 154.
[48] Matthew O Jackson "A crash course in implementation theory" (2001) 18 Soc. Choice Welfare 655 at 656.
[49] Jackson, above n 48, at 656.

actions that lead to the desired outcomes.[50] Consequently, this approach enables processes such as the negotiations for new international climate agreements to be designed in a way that will be more likely to lead to cooperative outcomes.[51]

The problem of achieving international cooperation on climate agreements, where parties have strong incentives to exaggerate their abatement costs and misrepresent their preferences, is also dealt with by implementation theory. Implementation theory is closely related to mechanism design.[52] Finding mechanisms that induce players to reveal their preferences is an important step to achieving collective climate action.[53] If countries are prevented from misrepresenting their abatement costs and environmental preferences (to negotiate weaker targets for themselves or reduce the likelihood of being committed to a more stringent target), significant climate action is more likely to occur.[54] Furthermore, punish and reward mechanisms to facilitate cooperation between states within international treaty making can also be designed.

Game theory models can also be a powerful tool of persuasion. Accurately presenting the game to players is likely to increase the chance of obtaining socially optimal outcomes.[55] When decisions makers at various levels of public policy and international diplomacy can view the game and see that cooperation is feasible by determining the possible outcomes, they will be better supported to negotiate more ambitious climate agreements. For example, in community games (where players do not make decisions independently), positive behaviour by one player can trigger reciprocal niceness by other players.[56] Players are incentivised to repay each other's kindness. In the climate context, a major emitter such as China or the US deciding to take unilateral voluntary action to reduce their emissions could

[50] Jackson, above n 48, at 656.
[51] Wood, above n 25, at 153.
[52] Wood, above n 25, at 165.
[53] Wood, above n 25, at 167.
[54] Wood, above n 25, at 167.
[55] Madani, above n 1, at 70.
[56] Madani, above n 1, at 70.

persuade other countries to do the same.[57] This increases the chance of greater trust developing between states and a socially optimal outcome being reached.[58]

The complex relationships underpinning issues such as carbon emissions, climate change, ecological and environmental damage, and the strategic options for actors to mitigate this become more transparent and achievable through the lens of game theory.[59] Game theory models play an essential role in encouraging climate mitigation action by major players. Ensuring incentives are set up right, rewarding cooperation between states and punishing free-riders, are all means to forward the negotiation of more ambitious climate agreements.[60] Using game theory, adequate strategies and recommendations can be adopted, supporting the likelihood of greater cooperation on international climate agreements.

C Limitations of game theory in climate change

Game theory, however, does have its limitations. Seen in a negative light, game theory models may threaten to discourage important action by major states, resulting in an even graver tragedy of the commons.[61] Scholars have cautioned the applicability of game theory models to climate change, noting significant limitations.[62] These include the set assumptions inherent to game theory and the lack of reliable information,[63] considerable uncertainty regarding probabilities and payoffs,[64] and overly simplistic game theory models and assumptions.[65] Madani warns us that the limitations of game theory models

---

[57] Madani, above n 1, at 70.
[58] Karen Pittel and Dirk T. G. Rubbelke "Transitions in the negotiations on climate change: from prisoner's dilemma to chicken and beyond" (2012) Int Environ Agreements 23 at 37; and Mond, above n 26, at 447.
[59] Kutasi, above n 32, at 55.
[60] James Dyke "Can game theory help solve the problem of climate change?" (13 April 2016) The Guardian <https://www.theguardian.com/science/blog/2016/apr/13/can-game-theory-help-solve-the-problem-of-climate-change>.
[61] Madani, above n 1, at 70.
[62] See generally Kutasi, above n 32, at 53; and Wood, above n 25, at 154; and Madani, above n 1.
[63] Wood, above n 25, at 154.
[64] Kutasi, above n 32.
[65] Thomas Dietz and Jinhua Zhao "Paths to climate cooperation" (2011) 108(38) Proceedings of the National Academy of Sciences of the United States of America 15671; and Madani, above n 1, at 71.

must be carefully considered when interpreting them or applying their results to policy, but some models are too simple to provide useful policy advice.[66]

Wood notes the importance of being aware of the limitations of game theory models when applying them to understand a problem.[67] Many applications of game theory require players to be rational, to have clear preferences and expectations of the unknown, and to act consistently with such preferences and expectations.[68] "The more accurate the available information, the better the decisionmaker will be able to predict the outcome associated with possible decisions, and thus make better choices."[69] However, the assumptions inherent to many game theory models do not always accord with reality, nor with experimental psychology.[70] Furthermore, the lack of reliable information regarding players' preferences and expectations makes it difficult to predict their strategy and the possible eventual outcome.[71] This has resulted in game theory not being widely used in studies on climate forecast problems, due to the increase in information and knowledge necessary to accurately develop and solve the games.[72]

Kutasi also discusses the dominant nature of uncertainty that is characteristic to climate change.[73] When attempting to model possible games, particularly the actions and reactions of different players, the characteristics of events, actions and behaviours must be known.[74] In the climate context, where considerable uncertainty exists, the payoff of the game is unknown.[75] It will likely depend on players' expectations and their reactions to other

---

[66] Above n 1, at 68.
[67] Above n 25, at 154.
[68] Wood, above n 25, at 154.
[69] Debra J Rubas, Harvey S J Hill and James W Mjelde "Economics and climate applications: exploring the frontier" (2006) 33(1) Climate Research 43 at 43.
[70] Wood, above n 25, at 154; and Elinor Ostrom "A Polycentric Approach for Coping with Climate Change" (Policy Research Working Paper no. WPS 5095, World Bank, 2009).
[71] Mond, above n 26, at 447.
[72] Rubas, Hill and Mjelde, above n 69, 47.
[73] Above n 32, at 44.
[74] Kutasi, above n 32, at 44.
[75] Kutasi, above n 32, at 44.

players' actions, or the likelihood of certain external events occurring.[76] Additional sources of uncertainty detailed by Kutasi include an asymmetry of information, lack of ability to carry out commitments, negative selection, decentralised governance, and short-sighted self-interest.[77] Such uncertainty, especially if they are multiplied, renders the likelihood of effective actions low. In the past, uncertainty has been used as a weapon in climate negotiations to keep opponents guessing, preventing them from taking quick and effective action.[78] Overall, the high number of players involved, different variables at play, and high level of uncertainty poses significant constraints to game theory modelling.

To complicate matters further, there is often a moral judgment at play when determining players' preferences.[79] For example, the inability to place an economic cost on carbon reflects that there is often a moral judgement at hand. Players in the climate context have been described as living "in different moral universes".[80] Countries often talk past each other. Their circumstances, especially when we compare the plight of rich developed nations to poor developing countries, can be as if worlds apart.

Some scholars' game theoretic characterisations of climate negotiations have also been critiqued as being overly simplistic, limiting the applicability of the models to the real world.[81] When considering DeCanio and Fremstad's "exhaustive treatment of the possible game-theoretic characterizations of climate negotiations between two players",[82] Madani identified numerous gaps in their analysis.[83] He focused on three key issues. First, that players were characterised incorrectly as making decisions independently without interaction with other players.[84] Second, that climate change games are not one-shot games,

---

[76] Kutasi, above n 32, at 44.
[77] Above n 32, at 45.
[78] Avinash Dixit and Susan Skeath *Games of Strategy* (2nd ed, W.W.Norton & Company, 2004) at 272.
[79] Mond, above n 26, at 446.
[80] Stephen J DeCanio and Anders Fremstad "Game theory and climate diplomacy" (2013) 85 Ecological Economics 177 at 183.
[81] See Madani's criticism of DeCanio and Fremstad's article, "Game theory and climate diplomacy" in Madani, above n 1.
[82] Above n 80, at 177.
[83] Above n 1.
[84] Above n 1, at 69.

where countries only choose between polluting and abating once, with no chance to switch strategies.[85] Third, that the "semi-static" game structure assumption did not accurately reflect the climate game as an evolving one.[86]

When applying game theory to complex dilemmas like climate change, there is a significant risk that models will be overly simplified.[87] Whilst Madani acknowledges that, "simplifications are integral to modelling complex conflicts", he asserts that "the effects of simplifying assumptions on the modelling outputs should not be overlooked when interpreting the results".[88] Failure to characterise games correctly and comprehensively can result in a failure to develop highly reliable policy conclusions.[89] In turn, this could result in more tragic climate outcomes. As Madani concludes:[90]

> In my opinion, prescribing policy actions that can affect the state of nature and the well-being of billions of people around the globe must not rely on simple game models that ignore some essential characteristics of the problem.

However, this is not all doom and gloom. Despite its limitations, game theory does have a key role to play in the climate context. It remains a useful tool to depict and address the complexities at play, particularly in climate agreement negotiations.

Some game theory concepts and models, like the Prisoner's Dilemma and the Tragedy of the Commons, rely on a distinct lack of information – particularly, the ignorance of others' intentions.[91] This can be taken to align with reality. Using implementation theory, different games can be designed to represent the climate negotiation context more accurately.[92]

[85] Above n 1, at 70.
[86] Above n 1, at 71-72.
[87] Dietz and Zhao, above n 65.
[88] Above n 1, at 68.
[89] Madani, above n 1, at 69.
[90] Above n 1, at 68.
[91] Mond, above n 26, at 446.
[92] Wood, above n 25, at 153.

Though players might not always have clear preferences and act rationally, game theory simply provides a method through which one can view the madness.

Furthermore, empirical work has shown that assumptions not aligning with reality can reflect positively on the climate cooperation crisis.[93] In situations such as the tragedy of the commons, scholars have shown that where there are suitable mechanisms that ensure individuals know what others are doing, shared resources can be used sustainably.[94] Additionally, when considering how human behaviour consideration relates to cooperation in the climate context, Ostrom found that a surprisingly large number of individuals facing collective action problems are likely to cooperate – much more than we would tend to expect.[95]

Thus, although game theory does suffer from some fundamental flaws, it is readily acknowledged that "all models are simplified representations of reality, full of limitations."[96] As some scholars have argued, "all models are wrong, but some are useful."[97] The practical question is therefore, how wrong must the model be to not be useful? This is the question the next part of this paper seeks to address.

IV Key Assumptions in Climate Negotiation Game Theory Models

When considering how game theory models can best support the negotiations of more ambitious climate agreements, we must pay regard to and address at least nine key issues and assumptions specific to the climate change problem.

As identified by Hsu, these include:[98]

---

[93] Elinor Ostrom, Amy R. Poteete and Marco A. Janssen *Working Together: Collective Action, the Commons and Multiple Methods in Action* (Princeton University Press, 2010); and Mond, above n 26, at 447.

[94] Elinor Ostrom *Governing the Commons, The Evolution of Institutions for Collective Action (Political Economy of Institutions and Decisions)* (Canto Classics, Cambridge University Press, 1991); and Ostrom, Poteete and Janssen, above n 93.

[95] Above n 70, at 10.

[96] Madani, above n 1, at 68.

[97] Box and Draper, above n 2, at 68.

[98] Above n 3, at 31-41.

1. The public good nature of greenhouse gas reductions
2. The free-rider effects of mitigation
3. Uncertainty regarding damages and adaptation costs
4. Discounting
5. The savings in mitigation costs of early mitigation
6. The role of technology
7. Mitigation costs are minimized by *coordinated* early action
8. International climate negotiations and mitigation actions take place over many time periods
9. International climate change negotiations defy traditional game-theoretic labels

First, it is readily acknowledged that the climate change problem is an example of a global public good.[99] Public goods must be non-excludable in provision and non-rival in consumption.[100] This means that once a good is provided it is necessarily provided to all, and that the enjoyment of a good by one person does not impact the enjoyment of the good by another.[101] Having less global warming by reducing greenhouse gas emissions is a *perfectly* non-excludable and non-rival activity.[102] It benefits everyone to reduce emissions, while everyone also has the incentive to see others take on the burden of reducing emissions.

Second, because of the first assumption, there is significant potential for parties to free-ride off other parties who choose to undertake costly climate mitigation actions. This is the classic collective action problem – if someone cannot be excluded from the use of a resource, they will not be motivated to contribute to the financing or sustaining of it but will instead free-ride off others' contributions[103]. This often takes the form of parties simply avoiding climate mitigation while others undertake it or avoiding costs attributable to the

---

[99] See Hsu, above n 3, at 31; and Wood, above n 25, at 153; and Ulrike Kornek and Ottmar Edenhofer "The Strategic Dimension of Financing Global Public Goods" (2020) 127 European Economic Review.
[100] Charles D Kolstad "Environmental Economics" (2011) 95 in Hsu, above n 3, at 31.
[101] Hsu, above n 3, at 31.
[102] Hsu, above n 3, at 31.
[103] Kutasi, above n 32, at 47.

research and development of technologies that will reduce greenhouse gas emissions.[104] The lack of any global coercive authority to enforce international agreements further exacerbates this problem.[105] This issue can be seen reflected in the Kyoto Protocol situation, where major emitters such as the United States and China effectively took a free-ride of the European Union and other developed countries' actions to mitigate climate change.[106]

Additionally, Hsu points out that a particularly troublesome aspect of the free-riding problem is that the incentives to freeride increase in response to the increase in mitigation undertaken by a country (or group of countries).[107] Cooperation on global warming is also linked to trade through a phenomenon known as carbon leakage.[108] This occurs when an increase in carbon dioxide emissions in one country results from the emissions reductions by another country with a stricter climate policy."[109] If a country chooses to unilaterally reduce their emissions, the reduced production can, in turn, increase the price of the good, which can lead to increased production of the good in another country that has *not* reduced its emissions.[110] This can lead to further economic benefits and an increase in emissions from the latter country, in effect, cancelling out some of the emissions reductions already achieved.

Though some scholars suggest the likelihood of free-riding can be prevented by punishing free-riders,[111] this is not possible in the carbon leakage situation. The only possible way to deter carbon leakage is to punish all players simultaneously, which would result in a

---

[104] Hsu, above n 3, at 32.
[105] Pittel and Rubbelke, above n 58, at 26.
[106] Pittel and Rubbelke, above n 58, at 26; and Lindsay Maizland "Global Climate Agreements: Successes and Failures" (17 November 2021) Council on Foreign Relations < https://www.cfr.org/backgrounder/paris-global-climate-change-agreements>.
[107] Above n 3, at 32.
[108] For more in-depth discussion read Dieter Helm, Cameron Hepburn and Giovanni Ruta "Trade, climate change, and the political game theory of border carbon adjustments" (2012) 28 Oxford Review of Economic Policy 368.
[109] Helm, Hepburn and Ruta, above n 108, at 376.
[110] Wood, above n 25, 159.
[111] See generally Heitzig, Lessmann and Zou, above n 36; and Caleiro, de Sousa, and de Oliveira, above n 46; and DeCanio and Fremstad, above n 80.

community loss.[112] Thus, by implication, to prevent free-riding any action to reduce emissions must be near-universal.[113] Furthermore, it is unlikely that major emitters such as the US and China can be effectively punished by other countries.[114] Late mitigating countries are likely to free-ride off the efforts of countries who undertake early mitigation.

Third, there is significant uncertainty regarding damages and adaptation costs.[115] Despite major developments in the past few decades, there is still scientific uncertainty about the effects, timelines, and the damages of climate change.[116] We do not know with certainty what effects will occur, when they will occur, where they may occur, how serious they might be, etc. There is still limited empirical experience on global warming, and the whole host of other unknowns that game theory modelists face.[117] Particularly in the face of climate change, uncertainty is accumulative.[118] This makes it increasingly difficult for nations to understand and communicate the risks to its citizens, and effectively plan for the future.[119]

Fourth, there is a time separation between countries undertaking the burden of mitigation and securing the benefits of avoiding climate change.[120] The further we are in time from experiencing the negative effects of climate change, the higher the levels of uncertainty about climate change and the lower the motivation of states to mitigate.[121] Yet, early mitigation is much cheaper and more efficient.[122] Conversely, the closer we get to realising the effects of climate change, the higher the certainty is, but the higher the mitigation costs are too. This is a timing dilemma the world finds itself in. With climate change, the most serious effects are likely to be felt by generations many years from now.[123] This time

[112] Dixit and Skeath, above n 78, at 49.
[113] Hsu, above n 3, at 32.
[114] Hsu, above n 3, at 33.
[115] Kutasi, above n 32; and Hsu, above n 3, at 33.
[116] Hsu, above n 3, at 42.
[117] Kutasi, above n 32, at 53.
[118] Kutasi, above n 32, at 44.
[119] Hsu, above n 3, at 34.
[120] Hsu, above n 3, at 34.
[121] Kutasi, above n 32, at 45.
[122] Hsu, above n 3, at 36.
[123] Wood, above n 25, at 161.

preference is incorporated into the idea of the 'discount rate', which is "the rate at which we discount future benefits in relation to present costs."[124] When developing a game-theoretic model, the discount rate should consider the countries' political behaviour and sensitivities regarding climate change, and that delayed action could result in irreversible ecological damage.[125]

Fifth, as discussed in the previous paragraph, early mitigation will be cheaper than delaying action. Once emissions rise above certain levels, cutting emissions will become more costly.[126]

Sixth, it is difficult for game-theoretic models to take technological innovation into account.[127] Development of technologies such as carbon capture and storage technology or geo-engineering measures could drastically change nations' ability to reduce emissions. However, models of technological learning rates do not assume large jumps in technological innovation, as they rarely occur.[128] Though the threat of climate change could result in a different outcome, it is unlikely that game-theoretic models will be able to accurately predict the role of technological innovation in climate mitigation. This also leaves aside questions of whether players would even have the necessary political capital to implement several technological solutions simultaneously.[129]

Seventh, early mitigation by one country is more costly than early mitigation *coordinated* between multiple countries.[130] If research and technological development costs can be shared between several countries, there is a greater potential of reaching research and development economies of scale.[131] Examples of early coordinated action include that of predominantly wealthy industrialised countries in deploying renewable energy

---

124 Mond, above n 26, at 446.
125 Hsu, above n 3, at 36.
126 Hsu, above n 3, at 36.
127 Hsu, above n 3, at 37.
128 Hsu, above n 3, at 36.
129 See generally Anthony Patt and others "Will policies to promote energy efficiency help or hinder achieving the 1.5°C climate target?" (2018) 12 Energy Efficiency 551; and Jessica Jewell and Aleh Cherp "On the political feasibility of climate change mitigation pathways: Is it too late to keep warming below 1.5°C?" (2020) 11(1) WIREs Climate Change 621.
130 Hsu, above n 3, at 39.
131 Hsu, above n 3, at 39.

technologies.[132] Now, renewable energy technologies are broadly competitive with current market energy prices.[133]

Eighth, many models of international climate negotiations are reduced to highly simplistic, static models which capture the collective action problem but not the intertemporal effects of actions.[134] Like Madani,[135] Hsu notes that the climate game is an evolving one.[136] In addition, Hsu proposes that a signalling mechanism could provide a way to recruit greater participation into multilateral and bilateral negotiations.[137]

Finally, Hsu argues that "the complexity of international climate negotiations is such that it is difficult to characterize using labels traditionally relied upon by game theoreticians to determine model structure".[138] He suggests that climate negotiations could be better modelled as dynamic games of incomplete information, and that a signalling model can be used to represent relationships between cooperating parties.[139]

In summary, Hsu argues that any model of international climate negotiations must address or at least be explicit about its assumptions regarding the nine issues we have discussed.[140] Failure to consider these issues when designing a game theory model of international climate negotiations will render the model of little use in conceptualising international climate negotiations.

---

132 Jewell and Cherp, above n 129, at 626.
133 International Renewable Energy Agency (IRENA) "Majority of New Renewables Undercut Cheapest Fossil Fuel on Cost" (press release, 22 June 2021); and Victoria Masterson "Renewables were the world's cheapest source of energy in 2020, new report shows" (5 July 2021) World Economic Forum < https://www.weforum.org/agenda/2021/07/renewables-cheapest-energy-source/#:~:text=Victoria%20Masterson&text=The%20cost%20of%20renewable%20technologies,plunged%2085%25%20in%20a%20decade.>.
134 Hsu, above n 3, at 40.
135 Above n 1, at 71.
136 Above n 3, at 41.
137 Above n 3, at 40.
138 Above n 3, at 40.
139 At 41.
140 At 31.

V Game Theory Models Applied to Climate Negotiations

This section introduces four types of game theory models and explains how they have been or can be applied to international climate negotiations. In discussing the game theory models their utility, applicability, and limitations to the climate negotiation context will be evaluated. This section will also consider the possible implications of accepting each model as accurately characterising the climate situation at hand.

A The Prisoner's Dilemma

Decision-making in situations of collective action can be represented by the well-known classic model of collective action known as the Prisoner's Dilemma. This game-theory construct reflects the tension between parties choosing to pursue individual selfish behaviour or the collective interest.

The classic model of collective action depicts the level of collective action as the sum of all individual players' contributions.[141] Individuals decide how much they wish to contribute to the collective action based on how much they will benefit from acting. This benefit is derived from the value of the collective action minus the cost of the individual player's contribution.[142]

Represented in its two-player form, the Prisoner's Dilemma involves two players who are unable to communicate having to choose between cooperating with the other player or not.[143] Each player's decision is made independently and simultaneously.[144] Each player has an incentive to act in their own self-interest, at the other player's expense. This is because the incentives and benefit of acting in their own individual self-interest is greater than choosing to cooperate. When both parties choose to act in their own self-interest, a less than optimal overall outcome is produced. If the parties had both chosen to cooperate,

---

141 Hale, above n 6, at 75.
142 Hale, above n 6, at 75.
143 The Investopedia Team "Prisoner's Dilemma" (3 September 2021) Investopedia <What Is the Prisoner's Dilemma and How Does It Work? (investopedia.com)>.
144 Wood, above n 25, 154.

the optimal outcome would have been produced. However, the incentives are such that players will always choose to not cooperate.[145]

Many scholars believe that climate change can be accurately represented by the Prisoner's Dilemma model.[146] The game-theoretic construct can be used as the basis for modelling situations such as whether countries should participate in international climate treaties, whether they should cheat on their emission reports, and how the short-term individual and long-term global interests meet regarding emissions.[147] Always present is the danger of free-riding and defection, which is the pre-eminent concern in collective action problems regarding public goods like climate change.[148]

There are a number of ways that the Prisoner's Dilemma can be conceptualised and understood in the climate context. First, we consider Wood's example of the Prisoner's Dilemma when seeking to achieve cooperation to reduce greenhouse gas emissions.[149] There are two players (countries) who can choose whether to *Pollute,* or *Abate.* All players will be collectively better off if emissions are reduced and both countries choose to *Abate* (*Abate, Abate)*. This is the social optimum, where the collective payoff is maximised. However, each player will be individually better off if they continue to *Pollute*, whilst the other player chooses to *Abate*. Each player prefers this situation (*Pollute, Abate),* because the country that plays *Pollute* will always be individually better off. If both countries choose to *Pollute (Pollute, Pollute),* both countries will be in a worse state than if they had chosen to cooperate *(Abate, Abate)*. All countries will be worse off when both choose to *Pollute* as emissions are not reduced. Herein lies the chief barrier to climate cooperation – countries will have a greater incentive to free-ride off others' efforts, than to choose to *Abate*. Hence, both players are far more likely to choose to *Pollute,* resulting in no climate cooperation, and all players being worse off.

---

[145] DeCanio and Fremstad, above n 80, at 180.
[146] See Caleiro, de Sousa, and de Oliveira, above n 46; and DeCanio and Fremstad, above n 80; and Pittel and Rübbelke, above n 58; and Kutasi, above n 32.
[147] Kutasi, above n 32.
[148] Aklin and Mildenberger, above n 4, at 8.
[149] Above n 25, at 155.

A second example is from DeCanio and Fremstad, who illustrate the Prisoner's Dilemma as a game between two Great Powers whose highest priority is to avoid a decline in their economic and military strength relative to the other.[150] Although both players would benefit from jointly reducing emissions (both countries playing *Abate*), when considering the costly price of abatement the worst outcome would be for one player to *Abate* while the other continues to *Pollute.* Negotiation of an enforceable international agreement to *Abate* would lead to the optimal solution. However, both countries would always have the incentive to defect, given their highest priority is prevailing in geopolitical competition. Neither country has an incentive to deviate from the Nash equilibrium (*Pollute, Pollute).* This explains how climate negotiations and policy making are fundamentally reciprocal. Players will not cooperate and implement the necessary policy reforms unless other players are certain to do so as well.[151] Unilateral cooperation is irrational when considering the economic costs to the individual player, as it implies that the player is willing to suffer the costs associated with others' free-riding.[152]

Similarly, Mond nicknames his conceptualisation of the Prisoner's Dilemma in climate negotiations as the "Polluter's Dilemma".[153] Imagining the two players as the US and China, he explains how the structure of the classic model will militate against a successful climate negotiation to limit emissions. If one side chooses to comply, the other side will gain economically by not complying, gaining a competitive advantage over the other. As such, there is always an incentive to defect. Additionally, once one player defects, the other will quickly follow suit. The chances of this occurring increases when the player who defects is a pivotal player, who is necessary to provide the global public good.[154] For example, a pivotal player would be a country like the United States or China, whose emissions make up a significant percentage of global CO2 emissions. Without pivotal

[150] Above n 80, at 180.
[151] Aklin and Mildenberger, above n 4, at 8.
[152] Aklin and Mildenberger, above n 4, at 8.
[153] Above n 26, at 448.
[154] Aklin and Mildenberger, above n 4, at 8.

players participating in international climate agreements, the chances of preventing significant climate change becomes far less likely to occur.[155]

The Kyoto Protocol has also been said to provide a good example of the Prisoner's Dilemma model playing out.[156] Players had little incentive to provide the public good in the shape of climate protection, as reducing emissions would have led to greater financial loss if other players did not follow suit.[157] Thus, it was safer for parties not to cut emissions. Additionally, the lack of binding obligations and sanctions on parties meant there was no guarantee that players would act in the way that they said they would.[158] The lack of mutual trust between parties increased players' incentive to free-ride on the abatement of other countries without contributing their own individual efforts to abate.[159] This was the dominant strategy of all parties. Hence, major emitters like the US and China effectively took a free-ride off other countries efforts to reduce emissions (such as the EU).[160]

Although there is strong support for the applicability of the Prisoner's Dilemma in the general climate context, it faces several limitations when considering the specific context of international climate negotiations.[161] This is predominantly because the model is overly simplistic in its choices and assumptions.[162] Countries do not face one simple choice of whether to pollute or abate, as in the examples given above. The climate situation gets far more complex when we consider that each player chooses *how much* pollution to emit, not simply whether to pollute or not, and that the damages from each player polluting is also dependent on the total amount of pollution emitted by all players.[163] The Prisoner's

---

[155] Aklin and Mildenberger, above n 4, at 8.
[156] Pittel and Rübbelke, above n , at 24; and Kutasi, above n 32; and Lila MacLellan "Game theory says the Paris Agreement looks like a winner for the climate" (17 April 2022) Quartz <https://qz.com/2099301/game-theory-says-the-paris-agreement-might-be-a-climate-winner/>.
[157] MacLellan, above n 156.
[158] Hsu, above n 3, at 22.
[159] MacLellan, above n 156.
[160] Pittel and Rübbelke, above n 58, at 26.
[161] See Hale, above n 6, at 76; and Aklin and Mildenberger, above n 4; and Madani, above n 1; and DeCanio and Fremstad, above n 80.
[162] Dietz and Zhao, above n 65; and Madani, above n 1, at 71.
[163] Wood, above n 25, at 156.

Dilemma situation also assumes that decisions are made independently, without communicating with other players.[164] However, in the climate negotiation context there is always communication between parties, increasing the likelihood of cooperation.[165] The nature of climate negotiations is such that parties often signal their intent to cooperate with each other and to form enforceable agreements.[166] Communication or signalling of this sort can work to lower early costs of mitigation and provide insight into player's intentions regarding future mitigation.[167] Yet, the prisoner's dilemma does not acknowledge this.

The normal form representation of the game also hides the constantly evolving and sequential nature of strategy and decision-making that occurs in international climate negotiations.[168] Framing the Prisoner's Dilemma as a one-shot game, without considering or providing for repeated play, leads to game models which do not reflect the reality of players changing decisions during the game, adopting mixed strategies, responding to actions by other players, and pursuing issue linkages.[169] Factors like issue linkages are crucial to expanding the feasible solution set and providing opportunities for players to make a "strategic loss" in some games to win in the overall linked game.[170] Isolating climate negotiation games from the wider international game context could result in misleading models and solutions being formed.[171] Furthermore, reducing climate negotiations to a simple, static prisoner's dilemma model fails to capture any intertemporal effects of action,[172] or the role of learning by playing a game repeatedly.[173]

Though DeCanio and Fremstad gave an example of the Prisoner's Dilemma in the climate context, they doubted whether their model accurately described the actual climate

[164] Wood, above n 25, at 156.
[165] Madani, above n 1, at 69.
[166] Mond, above n 26, at 446.
[167] Hsu, above n 3, at 40-41.
[168] Madani, above n 1, at 71.
[169] Madani, above n 1, at 70.
[170] Madani, above n 1, at 70.
[171] Madani, above n 1, at 70.
[172] Hsu, above n 3, at 40.
[173] Caleiro, de Sousa, and de Oliveira, above n 46, at 26.

negotiation situation.[174] Instead, they concluded that “there is no reason to assume that the Prisoner's Dilemma is the best description of the climate negotiations”.[175] One of the factors they considered was that major emitters, like the US or China, could be large enough that their business-as-usual emissions alone would disrupt the climate.[176] With some countries accounting for greater overall percentages of global emissions, a model that treats each country as identical and only a minor contributor to climate change is inapplicable.[177] Furthermore, a critical mass or number of countries would be needed to enforce a cooperative resolution.[178] This is not accounted for in the Prisoner’s Dilemma model. Thus, the Prisoner’s Dilemma model does not accurately represent the complexities of exchanges seen in international climate negotiations.

B The Chicken Game

The name of this game comes from the idea of two players driving towards each other head-on.[179] The one who swerves aside loses and is called a chicken. The winner is the one who does not swerve aside. However, if neither player swerves, then they will both crash. The worst possible payoff is when both players crash into each other. The best payoff occurs when one person ‘wins’ by driving straight, causing the other opponent to swerve and be the chicken. If both drivers choose to swerve, then neither player ‘loses’ honour, so this outcome is preferable to being the chicken.[180] There is no dominant strategy which prevails in this game.[181]

---

[174] Above n 80, at 185.
[175] At 185.
[176] At 184.
[177] At 184.
[178] At 184.
[179] DeCanio and Fremstad, above n 80, at 182.
[180] Janet Chen, Su-I Lu and Dan Vekhter “Chicken” <https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/index.html>.
[181] Pittel and Rubbelke, above n 58, at 24.

Assigning number values, the respective payoffs can be represented in the following payoff matrix table:[182]

	Swerve	Drive Straight
Swerve	2, 2	1, 3
Drive Straight	3, 1	0, 0

The Chicken game ('Chicken') belongs to the group of coordination games.[183] Unlike the Prisoner's Dilemma, mutual defection is not the worst outcome. The Chicken game also deviates from the Prisoner's Dilemma slightly by assigning different relative values to mutual inactivity and the unilateral provision of the public good.[184]

Several scholars have argued that the climate negotiation context can be modelled as an evolution from a Prisoner's Dilemma situation to the Chicken game.[185] This is due to the international climate negotiation game structure evolving over time, as parties try to coordinate action and change their behaviour, thereby changing the cost of different outcomes. Evolving from the Prisoner's Dilemma, the Chicken game has been said to apply where mutual defection of both players would bring about catastrophic climate consequences.[186] Hence, one party will eventually "chicken out" and take unilateral climate action. This will help them to avoid even higher costs in the future, due to the continual defection of both parties.[187] Yet, in case the opponent chooses to cooperate first, the beneficial strategy would be to take a free-ride.[188] Therefore, the dominant strategy is to *not* agree on mitigating emissions. Chickening out and taking unilateral action becomes a decision that players end up 'waiting out'.

[182] Chen, Lu and Vekhter, above n 180.
[183] Pittel and Rubbelke, above n 58, at 24.
[184] Pittel and Rubbelke, above n 58.
[185] Pittel and Rubbelke, above n 58; and Madani, above n 1, at 71.
[186] Pittel and Rubbelke, above n 58, at 26.
[187] Madani, above n 1, at 71.
[188] Pittel and Rubbelke, above n 58, at 24.

If the Chicken game does not describe the climate situation faced today, there is the chance that it will become the future model of international climate negotiations if all players keep defecting. The Chicken game threatens to represent the terminal structure of international climate negotiations.[189] This carries the risk that parties who are more affected by climate change will face higher costs and may be forced to solve the problem unilaterally.[190] Inherently, this assumes that negotiators *will* have the ability to prevent serious climate effects unilaterally.[191] This may not be the case in the future. However, in the short term, an understanding of the Chicken game might encourage players to take unilateral action before it becomes too late.

C Treaty Participation Game

The "Treaty Participation game" example referred to by Barrett,[192] is an example of an extensive form game which illustrates the use of the subgame perfect equilibrium.[193] We will consider the two-player version of this game, where the payoffs are equivalent to the Prisoner's Dilemma (discussed above). In this example taken from Wood's article, the game can be split into three stages:[194]

- Stage 1: All players simultaneously choose whether to be a signatory or a non-signatory.
- Stage 2: Signatories choose whether to play Abate or Pollute, with the objective of maximizing their collective payoff.
- Stage 3. Non-signatories choose simultaneously whether to play Abate or Pollute.

We use backward induction to determine the subgame perfect equilibrium (a strategy combination constituting a Nash equilibrium in every subgame of the entire game).[195]

[189] Madani, above n 1, at 72.
[190] Madani, above n 1, at 72.
[191] Pittel and Rubbelke, above n 58, at 27.
[192] Scott Barrett *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (Oxford University Press, Oxford Oxfordshire, 2003) in Wood, above n 25, at 158.
[193] Wood, above n 25, at 158.
[194] Above n 25, at 158.
[195] Wood, above n 25, at 157.

- At Stage 3: The Nash equilibrium (the position where no player can increase their payoff by changing their decision unilaterally once the outcome has been reached) is for players to choose *Pollute.* Hence, non-signatories will play *Pollute.*
- At Stage 2: If there is only one signatory, they will anticipate that the other player, a non-signatory, will play *Pollute.* Therefore, they will also play *Pollute.* However, if both countries are signatories, they will both choose to play *Abate*, as this will maximise their collective payoff.
- At Stage 1: If one country decides *not* to become a signatory, the other country will be indifferent to becoming a signatory. However, if one country decides to become a signatory, the other country will be better off by also becoming a signatory.

Therefore, the subgame perfect equilibrium has all countries becoming signatories.[196] The backward induction method of analysis used for extensive form games also gives us insights into how countries may choose between playing *Pollute* or *Abate*, depending on the actions of the other player.

When extending this game to include more than two players, only partial (not complete) cooperation will occur.[197] Barrett considered an agreement where players were assumed to have the same emissions costs and benefits, and signatories attempted to maximise their collective benefits while non-signatories tried to maximise their individual benefits.[198] He found that an agreement would be likely to be self-enforcing if no signatory could benefit from exiting the agreement, and no non-signatory could benefit from joining the agreement.[199] However, self-enforcing agreements were not likely to have full participation from players.[200] This leads to the conclusion that a significant barrier to negotiations on international climate treaties is that parties do not have incentives to participate or comply with agreements.[201] Countries are likely to drop out of international climate agreements.

---

[196] Wood, above n 25, at 159.
[197] Wood, above n 25, at 168.
[198] Above n 192, at 159.
[199] At 159.
[200] At 159.
[201] Wood, above n 25, at 159.

This is often seen.[202] Solutions to address this might include introducing reciprocal measures, side payments, issue linkage and trade sanctions.[203] These additional mechanisms which can support the likelihood of a cooperative outcome will be discussed in the next Part.

D Stag Hunt Game

Another possible game theory model that has been used to describe international climate negotiations is the Stag Hunt game (also known as the assurance or trust game).[204] This is a game where players do not have a motivation for free-riding.[205] In the game, the best option is for players to cooperate and shoot the stag instead of going after the lesser target, a hare. If they cooperate, the best outcome will be achieved. If they choose to go for the lesser target, it will take less time and effort, but the result will be worth far less than the stag. This makes players extremely cautious about acting in uncertain circumstances, often waiting to see how the other player(s) will act with before they decide how they will act.[206] Consequently, players are almost unable to seize a seemingly obvious opportunity to cooperate. After time passes, players tend to accept the second-best solution, which is to go after the lesser target.[207]

In the climate negotiation context, this game can be used to model a situation where less developed countries wait for larger countries to initiate climate action or choose to participate in and commit to a treaty.[208] It has been said that the 2015 Paris Agreement is structured like a stag hunt.[209] This is because signatories are only required to prepare and

---

[202] See generally about the US' withdrawal from the Kyoto treaty and Paris Agreement in Caleiro, de Sousa, and de Oliveira, above n 46, at 23; and Soila Apparicio and Natalie Sauer "Which countries have not ratified the Paris climate agreement?" (13 August 2020) Climate Home News <https://www.climatechangenews.com/2020/08/13/countries-yet-ratify-paris-agreement/>.
[203] Wood, above n 25, at 159.
[204] Kutasi, above n 32, at 49; and MacLellan, above n 156.
[205] Kutasi, above n 32, at 50.
[206] Kutasi, above n 32, at 50.
[207] Kutasi, above n 32, at 50.
[208] Kutasi, above n 32, at 50.
[209] MacLellan, above n 156.

report on their climate action strategies.[210] They are not obligated to meet certain targets, and there are no penalties.[211] It is up to each country to decide whether they are willing to give up their autonomy and cooperate with other countries to go after the stag. However, game theorists are doubtful that conceptualising the Paris Agreement as a Stag Hunt will encourage countries to choose to go after the stag.[212] Acting in one's own self-interest is more certain to lead to reliable, short-term payoffs.[213] As will be discussed further in the final part of this paper, domestic distributive conflicts and the domestic political environment are significant binding constraints on international climate policy action for most countries.[214] This could be an even more significant constraint on international climate negotiations than the collective action theory's incentive to free-ride.

VI Supporting the Development of an International Climate Negotiation Game

This section will discuss the factors that could aid the designing of a unique game that reflects international climate negotiations. It considers what this game should look like and what mechanisms can facilitate a cooperative outcome being reached.

A What should an international climate negotiation game look like?

Designing an international climate negotiation game must consider the reality of decision-making that occurs. Climate negotiations are a repeated process,[215] the climate negotiation game should reflect this. Negotiators can change decisions during the course of the game in response to other players' actions.[216] The value of repeated 'play' that enables mutual trust to be built between parties and provides opportunities to punish non-cooperating players must not be underestimated.[217] Therefore, unlike 'one-shot' games which have

---

[210] Paris Agreement Under the United Nations Framework Convention on Climate Change 54113 UNTS 3156 (opened for signature 16 February 2016, entered into force 4 November 2016), arts 4 and 13.
[211] Caleiro, de Sousa, and de Oliveira, above n 46, at 23.
[212] MacLellan, above n 156.
[213] MacLellan, above n 156.
[214] Aklin and Mildenberger, above n 4, at 92.
[215] Wood, above n 25, at 160.
[216] Madani, above n 1, at 70.
[217] Pittel and Rubbelke, above n 58, at 37.

been used to describe climate negotiations in simplistic models such as the Prisoner's Dilemma,[218] the international climate game is better suited to being a 'multi-shot' game where players can make multiple moves and counter-moves.[219] Instead of players only being given one choice to *Pollute* or *Abate* (as seen in one-shot games), multi-shot games allow players to switch strategies and counteract other players during the course of the game.[220] Additionally, the repetition of the game allows players the ability to select pure or mixed strategies, and to pursue issue linkages (encouraging players to strategically lose some games in order to win in the overall linked game).[221] The likelihood of a cooperative outcome being reached becomes considerably higher when considering international climate negotiations as a multi-shot game, rather than as a one-shot game.[222]

However, there are some reasons why the climate change context might not suit a repeated international climate negotiation game. First climate change itself is not a repeated game. The damages that arise from greenhouse gas emissions are dependent on cumulative emissions and are also more likely to be experienced in the future.[223] Furthermore, the outcome of global warming may be permanent. The idea that climate negotiations can be repeated in the same literal and metaphorical environment may not align with the Earth's reality. Second, repeated games still do not consider any difficulties which may arise with the interplay between international and domestic politics. Oftentimes, treaties can be negotiated but domestic ratification of the treaty occurs separately. Domestic interests and politics can work against international climate negotiations, especially if international treaty ratification is made difficult by domestic opposition.[224] These challenge the utility of a repeated game model.

---

[218] DeCanio and Fremstad, above n 80.
[219] Madani, above n 1, at 70.
[220] Madani, above n 1, at 70.
[221] Madani, above n 1, at 70.
[222] Wood, above n 25, at 160.
[223] Wood, above n 25, at 161.
[224] See generally Aklin and Mildenberger, above n 4.

The long-term perspective of international climate negotiations and the effect of climate change should also be incorporated into an international climate negotiation game. Literature has suggested that games in the climate context might be better modelled as *evolving*, not static or semi-static games.[225] Since many games are dependent on the availability of information – about factors such as climate change effects, time effects and the stock of emitted greenhouse gases in the atmosphere, an evolving game structure provides room for changing conditions to be considered.[226] Ignoring such constraints could lead to undesirable game structures. Conversely, acknowledging an evolving game structure early on might influence players' behaviour to reduce the risk of future losses and lower payoffs.[227]

B How to increase the likelihood of a cooperative outcome?

There are a number of mechanisms which could help to support the negotiations of climate agreements. These include pledges of voluntary contributions, conditional proposal mechanisms, self-financed contract enforcement solutions, signalling mechanisms and other incentives such as side payments, trade sanctions, issue linkages and moral stigma. We draw on implementation theory to discuss this.

The idea behind a voluntary contribution is that players pledge to make certain contributions for the collective good.[228] If the sum of the contributions is enough to provide the public good, then players will pay and provide that good.[229] If the total number of contributions is insufficient to provide that public good, then each player's contribution is refunded.[230] In the climate negotiation context, an example of a voluntary contribution can be found in the form of the conditional proposal mechanism, used when players' preferences lead to a prisoner's dilemma.[231] Countries will often state what they are prepared to contribute as part of the agreement in international climate negotiations. This

---

[225] Madani, above n 1; and Hsu, above n 3, at 40; and Pittel and Rubbelke, above n 58.
[226] Madani, above n 1, at 72.
[227] Madani, above n 1, at 71.
[228] Wood, above n 25, at 165.
[229] Wood, above n 25, at 165.
[230] Wood, above n 25, at 165.
[231] Wood, above n 25, at 166.

proposal may be confirmed and become a binding commitment later once the country knows other players' proposals.

A further conditionality that can occur during climate negotiations is when a country makes an unconditional commitment, but is willing to increase their commitment (e.g., emission reductions) dependent on other players' commitments.[232] An example of this is during the Copenhagen climate negotiations in 2009, where the European Union made an unconditional commitment to reduce its emissions by 20% compared to 1990 levels by 2020, but would look to further reduce its emissions to 30% if other countries were sufficiently committed.[233] Similarly, Australia made an unconditional commitment to reduce its emissions by 5%, and would increase that by up to 15% if certain commitments were met, and to a further 25% if additional conditions were met.[234] Although the Copenhagen climate negotiations were not incredibly successful,[235] it at least provides an example of how increasing commitments and a cooperative outcome can be supported based on conditionality. These mechanisms, however, do require that countries cannot backtrack from their conditional commitments – a problem of enforcement that remains to be addressed by major international climate agreements.

Self-financed contract-enforcement, or repeated games with imperfect monitoring, is also a potential solution to address common pool resource problems such as climate change.[236] This solution provides for users of the resource to watch and control each other. They are assumed to have complete information about the state of the resource and the behaviour of other users.[237] An example of this is found in the Kyoto Protocol, where players set emission quotas and monitored each other for control.[238]

---

[232] Wood, above n 25, at 166.
[233] Wood, above n 25, at 166.
[234] Wood, above n 25, at 166.
[235] Vaughan and Adam, above n 39.
[236] Kutasi, above n 32, at 47.
[237] Kutasi, above n 32, at 47.
[238] See generally the Kyoto Protocol to the United Nations Framework Convention on Climate Change 2303 UNTS 162 (opened for signature 16 March 1998, entered into force 16 February 2005).

Mechanisms that discourage defection and punish non-cooperation are also vital to support the likelihood of a cooperative outcome being reached.[239] Side payments that discourage defection can be arranged.[240] However, this mechanism has a major flaw. Dependent on the specific cost of the side payment, if every player knew that side payments were possible to defect, each party would have an immediate incentive to announce their intention to defect.[241] This is because information that may indicate a player is less likely to cooperate will make it less attractive for the other players to cooperate.[242] In turn, attempts to extract side payments might signal to other players that a country is unwilling to take climate action, thus creating potential for defection to snowball.[243]

Trade restrictions as a form of issue linkage is an alternative way to punish non-cooperation.[244] Treaties can be designed in a way that imposes trade sanctions on countries that choose not to join and comply, whilst trade advantages can be extended to those who participate. For example, tariffs could be imposed on goods such as fossil fuels that are exported to noncooperating countries. Linking trade with cooperation by mechanisms such as the imposition of tariffs could have a considerable influence on changing countries' incentives within climate negotiations, increasing the levels of participation in international climate treaties.

Pressure on players to comply can also take softer forms, such as submitting non-cooperating countries to pariah status, and subjecting them to shame and moral stigma if they do not conform to the preferences of the "world community".[245] Even more broadly, using the idea of issue linkage, cooperating parties might agree to work together on various non-climate related international issues, whilst non-cooperating parties are treated as rogue

---

[239] DeCanio and Fremstad, above n 80, at 180.
[240] DeCanio and Fremstad, above n 80, at 180.
[241] DeCanio and Fremstad, above n 80, at 180.
[242] Hsu, above n 3, at 71.
[243] Hsu, above n 3, at 129.
[244] Wood, above n 25, at 157.
[245] DeCanio and Fremstad, above n 80, at 181.

states.[246] Such consequences could influence countries' preferences and be important in their consideration of the consequences of not cooperating in climate agreements.

Hsu discusses the importance of a mechanism where countries can 'signal' to each other their willingness to mitigate climate change, which works to address the various uncertainties that encourage free-riding and non-cooperation.[247] Signalling future intentions enables parties to convey information to each other, which can induce a cooperative outcome. This type of mechanism can be seen in the Paris Agreement 2015.[248] Parties are required to "prepare, communicate and maintain a nationally determined contribution (NDC) and to pursue domestic measures to achieve them",[249] and are also invited to "formulate and submit long term low GHG emission development strategies (LT-LEDS)".[250] These measures improve transparency and communication between parties, reducing the likelihood of free-riding. It also encourages cooperation by reducing key uncertainties.

Overall, scholars have made a number of suggestions of factors that should be considered when analysing climate change negotiations. Madani suggests that analysis should be taken out of the overly simplistic 2 x 2 framework, more attention should be paid to the essential characteristics of climate change negotiations and negotiators, the heterogeneity of players' payoff functions and powers, and the "possibility of forming coalitions, issue linkage, strategic loss, counteraction, reward and punishment, cheap talks, and playing "community" games should be considered".[251] DeCanio and Fremstad suggest that practical policy will depend on factors such as how serious the long-term climate effects

[246] Above n 80, at 185.
[247] Above n 3, at 40 and 55.
[248] Paris Agreement, above n 210.
[249] "Key aspects of the Paris Agreement" UNFCCC <https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement>.
[250] "What is the Paris Agreement" UNFCCC <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement#:~:text=Its%20goal%20is%20to%20limit,neutral%20world%20by%20mid%2Dcentury.>.
[251] Above n 1, at 72.

are considered to be and whether governments of major emitters genuinely care about future generations.[252] Hsu concludes:[253]

> The complexity of the climate change problem is such that no single model can capture all of the strategic behaviors in play in international climate negotiations. What seems more useful is a simple model to highlight the most prevalent and important behaviors, and suggest some non-obvious policy directions.

Again, this brings us back to the question of, "how wrong does a game theory model have to be, to not be useful?" This author suggests that the game theory assumptions, models, and factors encouraging cooperation, as discussed above, do have utility. Although they might not provide a complete, accurate depiction of all the complexities and intricacies of international climate negotiations, this is not what is expected, nor required, of game theory models. After all, "all models are simplified representations of reality, full of limitations."[254] However, in the next Part of this paper we will consider the broader issue of whether the collective action lens we have used is the *best* way to consider international climate negotiations. We will reflect on alternative constraints which might be blocking a cooperative climate outcome being reached.

VII Is Collective Action the Best Lens to consider International Climate Negotiations?

This paper has shown the widespread convergence on collective action theory by scholars as the preferred lens to look through to explain the global failure to mitigate climate risks.[255] With the incentives such that everyone benefits from a stable climate, but each country faces better short-term prospects continuing their business-as-usual emissions rate,[256] the pre-eminent and central concern dominating all climate negotiation from the collective

---

[252] Above n 80, at 186.
[253] Above n 3, at 31.
[254] Madani, above n 1, at 68.
[255] Aklin and Mildenberger, above n 4, at 73; and Victor, above n 5.
[256] DeCanio and Fremstad, above n 80, at 181.

theorist's perspective is the problem of free-riding.[257] International climate treaties like the Kyoto Protocol and the Paris Agreement have sought to reduce the likelihood of free-riding by increasing transparency and certainty regarding countries' individual commitments, with the hope of increasing compliance.[258] However, this logic of collective action and the belief that free-riding constitutes the most significant constraint on effective climate action has been severely challenged.[259] Scholars have questioned whether collective action theorists have been misguided in using game theory to diagnose and present "the wrong solution to a misunderstood problem".[260] This final Part will address these concerns.

First, it considers whether the outcome of climate negotiations can be better explained by distinguishing it from climate policies. Then, it considers distributive politics as an alternative theoretical framework to collective action theory. Finally, it discusses Thomas Hale's proposed 'Catalytic Cooperation' model, a new form of cooperation which challenges the diagnosis of the collective action problem and provides a framework that combines domestic and international politics.[261]

A Collective Action vs Distributive Politics Account

The lack of empirical evidence to substantiate the collective action claim that climate policies are primarily structured by free-riding concerns, contrasts with the theory's prominence and pervasive influence.[262] The source of the climate problem being arguably misunderstood can be traced back to the influence of the success of the Montreal

---

[257] See generally William Nordhaus "Climate Clubs: Overcoming Free-Riding in International Climate Policy" (2015) 105(4) American Economic Review 1339; and Stefano Carattini, Simon Levin and Alessandro Tavoni "Cooperation in the Climate Commons" (working paper no 292, Centre for Climate Change Economics and Policy, 2017); and Kutasi, above n 32; and Mond, above n 26; and Pittel and Rubbelke, above n 58.

[258] See generally Melissa Denchak "Paris Climate Agreement: Everything You Need to Know" 19 February 2021 NRDC https://www.nrdc.org/stories/paris-climate-agreement-everything-you-need-know; and Aklin and Mildenberger, above n 4.

[259] Aklin and Mildenberger, above n 4, at 79.

[260] Aklin and Mildenberger, above n 4, at 79; and Victor, above n 5.

[261] Above n 6.

[262] Aklin and Mildenberger, above n 4.

Protocol,[263] which found its roots in traditional collective action theory.[264] Since then, the concern about free-riding has dominated climate negotiations. The failure of the Kyoto Protocol was blamed for the lack of mechanisms to prevent free-riding.[265] The Paris Agreement sought to remedy this by seeking wide-spread participation and voluntary cooperation to increase the likelihood of compliance.[266] This was done to address the monitoring problem,[267] to enable states to know whether other states would comply with their stated obligations or not.[268] However, drawing on quantitative and original qualitative data, some scholars have found it difficult to empirically substantiate this perspective in many climate politics cases.[269]

Aklin and Mildenberger suggest that the source of this mismatch between collective action theory and empirical evidence can be drawn to a conflation between the distinct ideas of climate *outcomes* and climate *policies*.[270] The difference between these distinct concepts is that climate policy takes its form from governments' preferences, as shaped by the views of their pro- or anti-climate reform constituencies, whereas climate outcomes can be characterised by negative externalities.[271] Consequently, divisions in the domestic political and economic stakeholders can be said to trigger distributive conflict over national climate policymaking. Through this lens of distributive conflict, an alternative explanation for the empirical patterns seen within international climate negotiations can be found. It is important to note, however, that the dynamics of domestic distributional conflict can differ from international conflict.[272] Our focus is simply on whether the collective action theory's

---

[263] Montreal Protocol on Substances that Deplete the Ozone Layer 1522 UNTS 26369 (opened for signature 16 September 1987, entered into force 1 January 1989).

[264] Aklin and Mildenberger, above n 4, at 95.

[265] Barrett, above n 192, at 155.

[266] Paris Agreement, above n 210, art 6.

[267] Kutasi, above n 32, at 49; and Michihiro Kandori "Repeated Games" in *The New Palgrave Dictionary of Economics* (Palgrave, Tokyo, 2006); and Aklin and Mildenberger, above n 4, at 81.

[268] Paris Agreement, above n 210, art 7.

[269] Aklin and Mildenberger, above n 4, at 75.

[270] At 78.

[271] Aklin and Mildenverger, above n 4, at 78; and DeCanio and Fremstad, above n 80, at 177; and Hsu, above n 3, at 25.

[272] Aklin and Mildenberger, above n 4, at 81.

pre-eminent concern about free-riding constitutes the most significant constraint on climate policy.

Distributional accounts emphasise the importance of domestic actors and the institutions in which they evolve.[273] Literature on distributive conflict in the climate context suggests two possible broad explanations.[274] First, that close ties between government policy makers and ‘special interest’ sectors can result in undue control being exerted over the policy-making process.[275] Second, that sectoral and ideological balance of power between climate policy opponents and proponents often determine climate policy outcomes.[276] Though the inability of institutions to combat free-riding has been mooted as the most significant constraint on international climate negotiations and consequently climate policy making, empirical evidence reveals that there has been extensive unilateral national climate policymaking, even in the absence of a binding international climate treaty.[277] This challenges the collective action theory hypothesis that countries will reciprocally adopt or defect from climate agreements, depending on what other players choose to do.[278] It instead supports the distributional politics view that climate policies are more likely to be affected by domestic actors and institutions.

For example, despite the defection of the US, a major emitter and player, from the Kyoto Protocol, the flagship Emissions Trading Scheme was still pushed through by the European

[273] See generally Mark Purdon “Advancing Comparative Climate Change Politics: Theory and Method” (2015) 15(3) Global Environmental Politics 1; and David Houle, Erick Lachapelle and Mark Purdon “Comparative Politics of Sub-federal Cap-and-Trade: Implementing the Western Climate Initiative” (2015) 15(3) Global Environmental Politics 48.
[274] Aklin and Mildenberger, above n 4, at 83.
[275] See an example of special interest control between carbon-intensive economic sectors and governmental actors in Ian Bailey and others "The fall (and rise) of carbon pricing in Australia: a political strategy analysis of the carbon pollution reduction scheme” (2012) 21(5) Environmental Politics 691; and further discussion in Aklin and Mildenberger, above n 4, at 80.
[276] Aklin and Mildenberger, above n 4, at 83; and Erick Lachapelle, Robert MacNeil and Matthew Patterson “The Political Economy of Decarbonisation” (2017) 22(3) New Political Economy 311.
[277] Aklin and Mildenberger, above n 4, at 85.
[278] See an example of this assumption in the context that early mitigation would be met with reciprocal mitigation, rather than free-riding in Hsu, above n 3, at 41.

Union in 2005.[279] Aklin and Mildenberger even suggest that the there was a greater adoption of unilateral climate policies in the intervening period between the US' rejection of the Kyoto Protocol and the failed efforts to negotiation a climate treaty in Copenhagen in 2009.[280] This evidence runs contrary to collective action theory, which would predict that the defection of the US from the Kyoto protocol (constituting a free-ride off other player's climate mitigation efforts) would result in other players likewise defecting from their climate commitments.[281] Other examples of national policy-making inconsistent with collective action theory predicting conditional cooperation, defection, and policy inaction where free-riding is likely to occur, include the emersion of national carbon taxes in countries like Norway and Finland prior to international agreements like the Kyoto Protocol. Many further examples of climate policies being enacted unilaterally can also be found.

Furthermore, Milkoreit's detailed study of global climate negotiators involved in the 2012 COP found that negotiators' understanding of the nature of the climate problem was often conditioned by their actor identity, including how they perceived costs and benefits associated with particular multilateral design architectures.[282] Major concerns about climate policy free-riding did not appear to be central to their mental models, or a serious cognitive constraint on their beliefs.[283]

Overall, these factors cause Aklin and Mildenberger to conclude that domestic distributive conflicts are a binding constraint on international climate policy action, perhaps far more than free-riding concerns.[284] Climate policy inaction can be explained by both collective action theory and distributive politics. The survival of governments depends largely on the preferences of their domestic constituencies, influenced by distributive conflict. However,

---

[279] "EU Emissions Trading Scheme (EU ETS)" European Commission <https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en>.

[280] Above n 4, at 83.

[281] See generally Mond, above n 26, at 441; and Caleiro, de Sousa, and de Oliveira, above n 46.

[282] Manjana Milkoreit *Mindmade Politics* (Cambridge, MA: MIT Press, 2017) in Aklin and Mildenberger, above n 4, at 89.

[283] Aklin and Mildenberger, above n 4, at 89.

[284] At 92.

global collective action models tend to be blind to the force of local politics, instead deeming states as the fundamental players.[285] Does this make the collective action theory models we have discussed up this point, pointless? No. Instead, like Aklin and Mildenberger, we conclude that free-riding concerns should be considered alongside other potential breaking points in a richer theoretical model that does not focus on a single cause.[286] For that, we turn to a discussion of Thomas Hale's "Catalytic Cooperation" model, which challenges the 'free-rider' diagnoses of the climate negotiation problem with implications for the institutional remedies.

B The Catalytic Cooperation Model

The 'Catalytic Cooperation' model proposed by Thomas Hale challenges three of the core assumptions underlying the prisoner's dilemma. These assumptions are that:[287]

1) Contributing to the collective good is costly to the contributor
2) Players have relatively symmetric preferences, oriented only towards prioritizing economic performance.
3) Others' actions will dissuade players from acting, and that players' preferences will not change over time.

Hale revises the assumptions above and proposes that the presence of three conditions in climate mitigation shifts the classic model of collective action to a 'catalytic' model of collective action. These conditions are the presence of joint products, preference heterogeneity, and increasing returns.[288] Hale's revised 'catalytic' model of collective action therefore proposes that: the benefit to a player which induces them to cooperate is the sum of the value derived from the collective action, minus any cost of action, plus any private benefit.[289] Due to the three revised conditions, the chief barrier to cooperation is no longer the threat of free-riding, but the lack of incentive to act in the first place.[290]

---

[285] Aklin and Mildenberger, above n 4, at 96.
[286] At 75.
[287] Above n 6.
[288] At 74.
[289] At 80-81.
[290] At 73.

The first condition suggested by Hale that changes the classic collective action model to a 'catalytic' collective action model, is the existence of private co-benefits, which are also known as joint products.[291] Climate policies provide direct primary benefits – those derived from pursuing the climate policy's primary aim, as well as ancillary benefits.[292] Ancillary benefits are the monetised secondary, side benefits or co-effects of climate policy.[293] These include effects such as the mitigation of air pollution, prevention of soil erosion and biodiversity loss, rise in employment and competitiveness levels, and technological and developmental impacts, as well as collective public benefits.[294] Cumulatively, the benefits represent the domestic public goods for individual countries.[295] The magnitude of ancillary benefits work to temper the potential cost of climate mitigation, impacting the cost-benefit analysis involved when states negotiate climate agreements. Consequently, the benefits of cooperating can incentivise those who may not hold strong pro-climate views and who would otherwise be reluctant, to act.[296] In certain situations, the benefits of acting can become the primary focus, and climate mitigation becomes the result. A high level of ancillary benefits can change a player's dominant strategy to that of mitigation.[297] This takes on particular importance in developing regions, where the existence of ancillary benefits increases willingness to engage in climate negotiations.[298] Industrialised countries stand to lose more than developing countries in the case of mutual defection compared to unilateral participation and tend to have greater relative gain compared to developing countries.[299] This difference in payoff makes it even more important to consider the

[291] At 76-77.

[292] Pittel and Rübbelke, above n 292.

[293] IPCC "Global, Regional, and National Costs and Ancillary Benefits" in *Climate Change 2001 – Mitigation* (Cambridge, Cambridge University Press, 2001); and Pittel and Rübbelke, above n 292; and Alan Krupnick, Dallas Burtraw and Anil Markandya "The Ancillary Benefit's and Costs of Climate Change Mitigation: A Conceptual Framework" (Conference Paper: Workshop on assessing the ancillary benefits and costs of greenhouse gas mitigation strategies, Washington, DC, USA, 2000) at 1.

[294] Pittel and Rubbelke, above n 58, at 30.

[295] Pittel and Rubbelke, above n 58, at 30.

[296] Fergus Green "Anti-fossil fuel norms" 150(1) Climatic Change 103.

[297] Pittel and Rubbelke, above n 58, at 32.

[298] Pittel and Rübbelke, above n 292.

[299] Pittel and Rübbelke, above n 292, at 216.

developing world, who tend to abstain from climate commitments, and who need to be incentivised to instead participate.[300]

An example of the presence of joint products positively influencing environmental outcomes, are the co-effects of gaining market shares in first-mover companies which successfully introduced substitutes for phosphates and CFCs, thus reducing emissions.[301] In Hale's "Catalytic Cooperation" model, these first-mover advantages can be viewed as a special form of ancillary benefits, which increases the payoff obtained by a country that chooses to mitigate while others do not.[302] These advantages tend to increase the attractiveness of a mitigation strategy.[303]

Hale's second condition is the existence of preference heterogeneity, which contrasts to the traditional Prisoner's Dilemma assumption that players' preferences are symmetrical, tending towards the prioritisation of economic performance.[304] The value players place on the costs and benefits involved in mitigation, and in mitigation itself, vary significantly. As discussed above, players preferences can include nonmaterial costs and benefits which are likely to vary by country, population density, and other factors.[305] For example, a country's adoption of a voluntary pledge to reduce emissions is dependent on its capacity to bear costs, and to act independently from private interests.[306] A government reliant on its fossil fuel sector might be severely hindered from pursuing climate policies that focus on energy transitions, compared to other states. Likewise, the climate situation, costs and benefits involved, and the value placed on collective action is likely to vary significantly when comparing a small island state to a major emitter such as China or India. Such variation in preferences requires careful and comprehensive analysis. This is being taken into consideration in Hale's model.

---

[300] Pittel and Rübbelke, above n 292, at 218.
[301] Pittel and Rubbelke, above n 58, at 33.
[302] Pittel and Rubbelke, above n 58, at 36.
[303] Pittel and Rubbelke, above n 58, at 37.
[304] Above n 6, at 77-78.
[305] Darby W Jack and Patrick L Kinney "Health co-benefits of climate mitigation in urban areas" (2010) 2(3) Current Opinion in Environmental Sustainability 172 at 175.
[306] Jewell and Cherp, above n 129, at 626.

The third condition Hale considers is the concept of increasing returns.[307] Under the traditional assumption of collective action theory, players' preferences do not change over time.[308] One player's actions will influence another's actions, simply due to the incentive to free-ride. Hale challenges this by considering the possibility of prior actions and policy choices being built upon and reinforcing each other, generating further action over time.[309] This is the concept of increasing returns. Not only may the material costs of action change (for example, the cost of renewable energy technologies steadily decreased as the technology was developed, becoming price-competitive with established products),[310] but the political processes of preference formation can also change.[311] Additionally, network and learning effects are likely to spur, accruing more value for users and stakeholders.[312] Where research costs can be shared amongst several countries, research and development economies of scale may be reached.[313] This increases the likelihood of effective climate mitigation occurring and provides greater hope that a critical mass can eventually be reached. If a critical mass can be reached, cooperative action may become progressively self-reinforcing. This can lead to a greater diffusion of norms and policies, and more ambitious climate agreements being made, over time.

Hale's model has significant implications for cooperation. First, the combination of the conditions of joint products and preference heterogeneity is likely to generate more unilateral action and increase the likelihood of cooperation becoming the dominant strategy.[314] This is because a low level of action might produce a sufficient level of benefit to outweigh the costs of acting. As time progresses, climate action becomes more desirable

---

[307] Above n 6, at 78-80.
[308] Wood, above n 25, at 154.
[309] Above n 6, at 78.
[310] IRENA, above n 133; and Manfred Fischedick and others "Mitigation Potential and Costs" in Ottomar Edenhofer and others (eds) *Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change* (Cambridge, Cambridge University Press, 2011) 791.
[311] Hale, above n 6, at 79.
[312] Hale, above n 6, at 79-80.
[313] Hsu, above n 3, at 56.
[314] Hale, above n 6, at 82-83.

as the consequences of not cooperating becomes riskier. Taking into consideration the ancillary and collective benefits may tip the scales towards cooperation. Second, if increasing returns are of sufficient strength, a 'tipping' point' where action becomes strictly preferable and self-reinforcing may be reached.[315] This will be encouraged by the reduction in perceived costs and increased benefits as discussed above. Players' preferences are also likely to alter consequent to their and other's actions. Third, an ever-increasing number of players cooperating will induce more players to cooperate.[316] Scholars have suggested that "when groups are heterogeneous and a good has high jointness of supply, a larger … group can have a smaller critical mass."[317] Following Hale's 'catalytic cooperation' logic, increasing the number of players will increase the likelihood that there will be at least a few players who have pro-climate action preferences.[318] The changed distribution of preferences will increase the likelihood of one or more players having a sufficiently strong preference to act, hopefully setting off additional chain reactions.

The above assumptions detail how the 'Catalytic Cooperation' model can provide a pathway to greater cooperation in international climate negotiation. However, Hale's model may still be insufficient to address a number of political and practical concerns surrounding climate negotiations. Even considering the changes made by Hale's model, it is still unlikely that the necessary changes can be induced to avoid overshooting the 1.5C° target set by the 2015 Paris Agreement.[319] Although there are significant ancillary benefits to pro-climate action policies, these may still be outweighed when considering the political costs involved.[320] This is particularly significant when climate negotiations are considered through the lens of distributive conflict. It may be politically feasible to deploy one climate solution, however, a multitude of changes are required. It may not be politically feasible to

---

[315] Hale, above n 6, at 84.
[316] Hale, above n 6, at 84-85.
[317] Gerald Marwell and Pamela Oliver *The Critical Mass in Collective Action* (Cambridge, UK, Cambridge University Press, 1993) in Hale, above n 6, at 84.
[318] At 91.
[319] Catharina Lee "Catalytic Cooperation: An incentive or disincentive to collective climate action?" (LLB (Hons) Seminar paper, Victoria University of Wellington, 2022).
[320] Thomas Bernauer and Liam F. McGrath "Simply reframing unlikely to boost public support for climate policy" (2016) 6 Nature Climate Change 680.

deploy *multiple* climate solutions at once,[321] especially when considering domestic constraints. Implementing several solutions simultaneously may spread political capital too thin, diminishing the likely success of all.[322] Additionally, the concept of increasing returns might work against climate action and cause pro-carbon policies to become entrenched.[323]

The change in assumptions in Hale's "Catalytic Cooperation" model can arguably be used in favour of both a pro- and anti-climate view. However, Hale's model better reflects the reality at hand when considering international climate negotiations, in comparison to the other traditional collective action game theory models we have considered. Catalytic Cooperation is essential when considering ways that negotiations of more ambitious climate agreements can be supported. The more accurate characterisation and account of climate negotiations and the consideration of broader sources of constraints in Hale's model can facilitate greater cooperation. Though this model may not perfectly depict the climate negotiation situation, it is still a useful tool to analyse, persuade, and incentivise action amongst key actors.[324] This may tip the scales in favour of greater climate action.

VIII Conclusion

This paper has introduced the concept of game theory in the international climate negotiation game. It has discussed how game theory is a useful tool to help us understand the incentives behind strategies adopted by parties in international climate negotiations and shows how game theory can be used to explain the resultant outcomes. Four traditional game theory models, their applicability to international climate negotiations, and the assumptions that must be addressed to accurately design an international climate negotiation game have been considered. The key constraints to game theory in international climate negotiations generally, and specific to the traditional game theory models discussed, were also outlined.

[321] Patt and others, above n 129; and Jewell and Cherp, above n 129, at 627.
[322] Jewell and Cherp, above n 129, at 627.
[323] Jewell and Cherp, above n 129, at 625.
[324] Madani, above n 1, at 72.

Ultimately, game theory models are not without their limitations. The climate context is fundamentally uncertain. Assumptions must be made, and whether these are to a sufficient degree to render game theory models useless in this context, is a crucial question that is left up for debate. Following the logic that “all models are simplified representations of reality, full of limitations”,[325] game theory in international climate negotiations can be treated no differently. No single model can capture all the strategic behaviours at play in international climate negotiations.[326] A model must not be overly simplistic,[327] yet should be simple enough to highlight the most prevalent and important behaviours, without detracting from the key scenario at hand.[328] In this sense, game theory models are analogous to a “Goldilocks and the Three Bears” type situation – in most cases, the porridge will be either too hot, or not hot enough. The more critical issue then becomes distinguishing between what models are better at characterising the international climate negotiation game, or worse.

As Aklin and Mildenberger suggest, climate policy inaction can often be explained equally by both collective action theory and distributive politics accounts.[329] Rather than vetoing game theory models completely, a good game theory model should consider the various domestic, political, and collective action constraints we have discussed. Thomas Hale’s ‘Catalytic Co-operation’ model of collective action considers both collective action theory and distributive politics as contributing sources to the collective action problem of non-cooperation, without losing sight of the traditional collective action model as its foundation. Catalytic Cooperation bridges the gap in analysis left by traditional collective action models and provides a new, ‘catalytic’ way forward.

By gaining insight into the root of the climate negotiation problem, we can design mechanisms and processes that will facilitate greater cooperation to solve that problem. Implementation theory addresses this by asking how a process can be designed to lead to

[325] Madani, above n 1, at 68.

[326] Hsu, above n 3, at 48.

[327] Madani, above n 1.

[328] Osborne, above n 10, at 48.

[329] Aklin and Mildenberger, above n 4, at 93.

the desired outcome.[330] The establishment of fair and efficient mechanisms that reward cooperation and punish non-cooperation,[331] such as trade sanctions and side payments,[332] are bound to have a key role to play in incentivising cooperation. If implementation theory can be utilised, mechanisms supporting cooperation can be designed, and key players can be convinced of the game they are in,[333] cooperation will become more feasible and better climate outcomes are likely to occur. The game theoretic investigation of cooperative mechanisms in the climate negotiation game may facilitate greater cooperation, ultimately supporting the negotiation of more ambitious climate agreements in the future.[334]

[330] Wood, above n 25, at 164.
[331] Heitzig, Lessmann and Zou, above n 36, at 15739.
[332] See generally DeCanio and Fremstad, above n 80.
[333] Kutasi, above n 32, at 55.
[334] Wood, above n 25, at 154.

IX Word Count

The text of this paper (excluding table of contents, footnotes, and bibliography) comprises approximately 12,435 words.

X Bibliography

A Books and Chapters in Books

António Bento Caleiro, Miguel Rocha de Sousa, and Ingo Andrade de Oliveira "Global Development and Climate Change: A Game Theory Approach" in Tiago Sequeira and Liliana Reis (eds) *Climate Change and Global Development: Contributions to Economics* (Springer, Cham, 2019).

Avinash Dixit and Susan Skeath *Games of Strategy* (2^{nd} ed, W.W.Norton & Company, 2004).

David G. Victor *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet* (Cambridge University Press, Cambridge, UK, 2011).

David Mond "Game theory and climate change" in *The Impacts of Climate Change* (Elsevier, 2021).

Elinor Ostrom *Governing the Commons, The Evolution of Institutions for Collective Action (Political Economy of Institutions and Decisions)* (Canto Classics, Cambridge University Press, 1991).

Elinor Ostrom, Amy R. Poteete and Marco A. Janssen *Working Together: Collective Action, the Commons and Multiple Methods in Action* (Princeton University Press, 2010).

Gerald Marwell and Pamela Oliver *The Critical Mass in Collective Action* (Cambridge, UK, Cambridge University Press, 1993).

IPCC "Global, Regional, and National Costs and Ancillary Benefits" in *Climate Change 2001 – Mitigation* (Cambridge, Cambridge University Press, 2001).

Mancur Olson *The logic of collective action: public goods and the theory of groups* (Harvard University Press, 1965).

Manfred Fischedick and others "Mitigation Potential and Costs" in Ottomar Edenhofer and others (eds) *Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change* (Cambridge, Cambridge University Press, 2011) 791.

Manjana Milkoreit *Mindmade Politics* (Cambridge, MA: MIT Press, 2017).

Martin J. Osborne *An Introduction to Game Theory* (Oxford University Press, 2003).

Michihiro Kandori "Repeated Games" in *The New Palgrave Dictionary of Economics* (Palgrave, Tokyo, 2006).

Scott Barrett *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (Oxford University Press, Oxford Oxfordshire, 2003).

Steven Tadelis *Game Theory: An introduction* (Princeton University Press, United Kingdom, 2013).

B Journal Articles

Anthony Patt and others "Will policies to promote energy efficiency help or hinder achieving the 1.5°C climate target?" (2018) 12 Energy Efficiency 551.

Colin F. Camerer "Progress in Behavioral Game Theory" (1997) 11 Journal of Economic Perspectives 167.

Darby W Jack and Patrick L Kinney "Health co-benefits of climate mitigation in urban areas" (2010) 2(3) Current Opinion in Environmental Sustainability 172.

David Houle, Erick Lachapelle and Mark Purdon “Comparative Politics of Sub-federal Cap-and-Trade: Implementing the Western Climate Initiative” (2015) 15(3) Global Environmental Politics 48.

Debra J Rubas, Harvey S J Hill and James W Mjelde “Economics and climate applications: exploring the frontier” (2006) 33(1) Climate Research 43.

Dieter Helm, Cameron Hepburn and Giovanni Ruta “Trade, climate change, and the political game theory of border carbon adjustments” (2012) 28 Oxford Review of Economic Policy 368.

Erick Lachapelle, Robert MacNeil and Matthew Patterson “The Political Economy of Decarbonisation” (2017) 22(3) New Political Economy 311.

Fergus Green “Anti-fossil fuel norms” 150(1) Climatic Change 103.

Garret Hardin “The Tragedy of the Commons” (1968) 162 Science 1243.

Gábor Kutasi “Climate Change in Game Theory Context” (2012) 13(1) Interdisciplinary Environmental Review 42.

George E.P Box and Norman R. Draper “Empirical model-building and response surfaces” (1987) John Wiley & Sons.

Ian Bailey and others "The fall (and rise) of carbon pricing in Australia: a political strategy analysis of the carbon pollution reduction scheme” (2012) 21(5) Environmental Politics 691.

Jessica Jewell and Aleh Cherp "On the political feasibility of climate change mitigation pathways: Is it too late to keep warming below 1.5°C?" (2020) 11(1) WIREs Climate Change 621.

Jobst Heitzig, Kai Lessmann and Yong Zou “Self-enforcing strategies to deter free-riding in the climate change mitigation game and other repeated public good games” (2011) 108 Proceedings of the National Academy of Sciences.

Karen Pittel and Dirk T. G. Rübbelke “Climate policy and ancillary benefits—A survey and integration into the modelling of international negotiations on climate change” (2008) 68 Ecological Economics 210.

Karen Pittel and Dirk T. G. Rübbelke “Transitions in the negotiations on climate change: from prisoner’s dilemma to chicken and beyond” (2012) Int Environ Agreements 23.

Karl-Henrik Robert and Goran Broman "Prisoners' dilemma misleads business and policy making" (2017) 140 Journal of Cleaner Production 10-16.

Kaveh Madani “Modeling international climate change negotiations more responsibly: Can highly simplified game theory models provide reliable policy insights?” (2013) 90 Ecological Economics 68.

Mark Purdon “Advancing Comparative Climate Change Politics: Theory and Method” (2015) 15(3) Global Environmental Politics 1.

Matthew O Jackson “A crash course in implementation theory” (2001) 18 Social Choice Welfare 655.

Michaël Aklin and Matto Mildenberger “Prisoners of the Wrong Dilemma: Why Distributive Conflict, Not Collective Action, Characterizes the Politics of Climate Change” (2020) 20(4) Global Environmental Politics 73.

Peter John Wood "Climate change and game theory" (2011) 1219 Ann. N.Y. Acad.Sci 153-170.

Shi-Ling Hsu "A Game Theoretic Model of International Climate Change negotiations" (2010) 19 NYU Environmental Law Journal 14.

Stephen J DeCanio and Anders Fremstad "Game theory and climate diplomacy" (2013) 85 Ecological Economics 177.

Thomas Bernauer and Liam F. McGrath "Simply reframing unlikely to boost public support for climate policy" (2016) 6 Nature Climate Change 680.

Thomas Dietz and Jinhua Zhao "Paths to climate cooperation" (2011) 108(38) Proceedings of the National Academy of Sciences of the United States of America 15671.

Thomas Hale "Catalytic Cooperation" (2020) 20(4) Global Environmental Politics 73.

Ulrike Kornek and Ottmar Edenhofer "The Strategic Dimension of Financing Global Public Goods" (2020) 127 European Economic Review.

William Nordhaus "Climate Clubs: Overcoming Free-Riding in International Climate Policy" (2015) 105(4) American Economic Review 1339.

C Reports

Summary of the Copenhagen Climate Change Conference: 7-19 December 2009 (International Institute for Sustainable Development: Earth Negotiations Bulletin, December 2009).

D Dissertations and Papers

Alan Krupnick, Dallas Burtraw and Anil Markandya "The Ancillary Benefit's and Costs of Climate Change Mitigation: A Conceptual Framework" (Conference Paper: Workshop on assessing the ancillary benefits and costs of greenhouse gas mitigation strategies, Washington, DC, USA, 2000).

Brian Spak “The Success of the Copenhagen Accord and The Failure of the Copenhagen Conference” (Substantial Research Paper, 2010).

Catharina Lee “Catalytic Cooperation: An incentive or disincentive to collective climate action?” (LLB (Hons) Seminar paper, Victoria University of Wellington, 2022).

Elinor Ostrom “A Polycentric Approach for Coping with Climate Change” (Policy Research Working Paper no. WPS 5095, World Bank, 2009).

Stefano Carattini, Simon Levin and Alessandro Tavoni “Cooperation in the Climate Commons” (working paper no 292, Centre for Climate Change Economics and Policy, 2017).

E International Treaties

Kyoto Protocol to the United Nations Framework Convention on Climate Change 2303 UNTS 162 (opened for signature 16 March 1998, entered into force 16 February 2005).

Montreal Protocol on Substances that Deplete the Ozone Layer 1522 UNTS 26369 (opened for signature 16 September 1987, entered into force 1 January 1989).

Paris Agreement Under the United Nations Framework Convention on Climate Change 54113 UNTS 3156 (opened for signature 16 February 2016), entered into force 4 November 2016).

F Internet Resources

Adam Hayes “Game Theory: Definition, Role in Economics, and Examples” (2 February 2022) Investopedia <https://www.investopedia.com/terms/g/gametheory.asp>.

Adam Vaughan and David Adam “Copenhagen climate deal: Spectacular failure – or a few important steps?” (22 December 2009) The Guardian <https://www.theguardian.com/environment/2009/dec/22/copenhagen-climate-deal-expert-view>.

"EU Emissions Trading Scheme (EU ETS)" European Commission <https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en>.

"Game Theory" (8 March 2019) Stanford Encyclopedia of Philosophy <https://plato.stanford.edu/entries/game-theory/index.html>.

International Renewable Energy Agency (IRENA) "Majority of New Renewables Undercut Cheapest Fossil Fuel on Cost" (press release, 22 June 2021).

James Dyke "Can game theory help solve the problem of climate change?" (13 April 2016) The Guardian <https://www.theguardian.com/science/blog/2016/apr/13/can-game-theory-help-solve-the-problem-of-climate-change>.

Janet Chen, Su-I Lu and Dan Vekhter "Chicken" <https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/index.html>.

"Key aspects of the Paris Agreement" UNFCCC <https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement>.

Lindsay Maizland "Global Climate Agreements: Successes and Failures" (17 November 2021) Council on Foreign Relations < https://www.cfr.org/backgrounder/paris-global-climate-change-agreements>.

Lila MacLellan "Game theory says the Paris Agreement looks like a winner for the climate" (17 April 2022) Quartz <https://qz.com/2099301/game-theory-says-the-paris-agreement-might-be-a-climate-winner/>.

Melissa Denchak “Paris Climate Agreement: Everything You Need to Know” 19 February 2021 NRDC <https://www.nrdc.org/stories/paris-climate-agreement-everything-you-need-know>.

Soila Apparicio and Natalie Sauer “Which countries have not ratified the Paris climate agreement?” (13 August 2020) Climate Home News <https://www.climatechangenews.com/2020/08/13/countries-yet-ratify-paris-agreement/>.

The Investopedia Team “Prisoner’s Dilemma” (3 September 2021) Investopedia <What Is the Prisoner's Dilemma and How Does It Work? (investopedia.com)>.

Victoria Masterson “Renewables were the world’s cheapest source of energy in 2020, new report shows” (5 July 2021) World Economic Forum <https://www.weforum.org/agenda/2021/07/renewables-cheapest-energy-source/#:~:text=Victoria%20Masterson&text=The%20cost%20of%20renewable%20technologies,plunged%2085%25%20in%20a%20decade.>.

“What is the Paris Agreement” UNFCCC <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement#:~:text=Its%20goal%20is%20to%20limit,neutral%20world%20by%20mid%2Dcentury.>.