UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE

U.F.R. DE SCIENCES ÉCONOMIQUES

Thèse pour le doctorat de Sciences Economiques
soutenue publiquement par

Lorenzo CERDA PLANAS

le 2 juillet 2015

TOWARDS GREENER SOCIETIES: NUDGING BEHAVIOUR AT A COUNTRY AND GLOBAL SCALE

Directeur de thèse : BERTRAND WIGNIOLLE

JURY

Alain AYONG LE KAMA	Professeur à l'Université Paris Ouest *(rapporteur)*
Scott BARRETT	Professeur à Columbia University *(rapporteur)*
Eugenio FIGUEROA	Professeur à l'Universidad de Chile
Guillaume HOLLARD	Directeur de recherche au CNRS
Katheline SCHUBERT	Professeur à l'Université Paris 1 Panthéon-Sorbonne

L'université Paris I Panthéon-Sorbonne n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

To my family

to those who are not with us any more

to my parents: Victoria and Félix

to Genoveva and Luis Alberto

Acknowledgments

Foremost, I am greatly thankful for Bertrand Wigniolle for accepting me as a PhD candidate and supervising my thesis during these years. He has been a great and constant support, helping my process with his kindness, wisdom and sharpness, despite his charged schedule. He has been of great help in making my engineer's mind seize the economic thought. It has been a great pleasure for me to work with him.

I cannot continue without thanking Eugenio Figueroa and Scott Barrett. Eugenio Figue-roa, from Universidad de Chile, has always been available for questions and discussions, the large distance that separates Europe to Chile never being a problem. He has received me in Chile through all these years and he has made the Facultad de Economía y Negocios (FEN) my second home. In the same manner, Scott Barrett has been extremely welcoming, first by sponsoring my exchange application to Columbia University twice, and then by making himself always available for discussion, while keeping up to his frenetic schedule. My stays in Chile and New York City would not have been this great and fruitful, if it had not be for these two people.

I am also deeply grateful to the members of the jury, Alain Ayong Le Kama, Scott Barrett, Eugenio Figueroa, Guillaume Hollard and Katheline Schubert. They have taken their time and effort in reading, analyzing and discussing my work thoroughly. They have been a great source of discussion and inspiration, which has helped me for improving my final dissertation. I am doubly thankful to Katheline, who not only was part of the jury, but she has also been helping me from the very beginning of this journey, back in my master dissertation and then in the bilan 18.

I thank Dominique Guegan, Hyppolite d'Albis and Mouez Fodha, whom as directors of the doctoral school have helped me in pursuing this research adventure. I am also thankful to professors that made their time to discuss with me along these years: Lisa Anouliès, Michaël Assous, Mireille Chiroleu-Assouline, Antoine d'Autume, Jean-Pierre Drugeon, Fanny Henriet, Katrin Millock and Hélène Ollivier.

In the same line, I thank the researchers that made my stays abroad fruitful and pleasant. In Chile: Juan Pablo Torres-Martínez, Rodrigo Wagner, Carlos Chávez, Wal-

ter Gómez, Jorge Dresdner, Felipe Vasquez and Juan Pablo Montero. In the United States: Michael Oppenheimer, Alexander Ovodenko, Thomas Sterner and Gernot Wagner. I also want to thank Georges Zaccour, who warmly invited to present in Montreal, Canada.

My doctoral life was made pleasant and easy due to the work and positive attitude of the staff of this faculty. Special thanks to Viviane Makougni, who was always available, helpful and caring. I am also grateful to Carmen Tudor, who has taken Viviane's place. I thank Elda Andre, Loic Sorel, Stephane, Rachad, Rachid and Joel for being always co-operative and for making the MSE an agreeable place to work. I am also deeply thankful to the library staff, which has always been extremely helpful and attentive. I would also like to thank their Chilean and American counterparts: Gema Menares, who made my stay in Chile and easy and pleasant as possible, and Valentina Paredes who arranged my presentations at FEN. To the Alliance program staff: Alessia Lefebure, Lauranne Bardin and Yvette Jusseaume.

I am also grateful to my PhD student colleagues, here and abroad, who made my graduate studies as pleasant as possible. To my office fellows and close friends: Alessandra, Baris, Diane, Elisa, Hamzeh, Mathias and Stefanija; and to those who have shared an office with me: Anastasia, Camille, Emmanuelle, Francesco and Sheng. To my colleges at the MSE: Ana, Anil, Anna, Antoine, Axelle, Basak, Can, Charlotte, Claire, Cléo, Diana, Djamel, Elliot, Esther, Esther, Guillaume, Ingrid, Lorenzo, Marie-Laure, Marion, Mehdi, Moutaz, Thais, Thomas, Victoire, Vincent and Yvan. To the people who were more than welcoming in New York City: Ana, Anthony, Beth, Cristian, Denyse, Dilnoor, Eugenie, Eyal, James, Kayleigh, María Cecilia, Marina, Prabhat, Semee, Stephanie and Vania.

Special thanks to Maria Kuecken, Kalyan Krishnamani and Federico Manolio, who helped me with different aspects of my papers. I could not have started pursuing my thesis without the support of two old friends in Chile: Juan Enrique Coeymans and Víctor Cortés. I am deeply thankful to them.

Last but not least, to my family and close friends back in Chile, for being encouraging and supporting all along this way. For those who made this long distance to look as a short one.

Avertissement

Mis à part l'introduction et la conclusion de cette thèse, les différents chapitres sont issus d'articles de recherche rédigés en anglais et dont la structure est autonome. Par conséquent, des termes "papier" ou "article" y font référence, et certaines informations, notamment la littérature, sont répétées d'un chapitre à l'autre.

Notice

Except the general introduction and the conclusion, all chapters of this thesis are self-containing research articles. Consequently, terms "paper" or "article" are frequently used. Moreover, some explanations, like corresponding literature, are repeated in different places of the thesis.

Contents

List of Figures

Introduction générale

Au cours des dernières années, la question de l'environnement prend de l'importance. A mesure que la population mondiale, ainsi que la consommation d'énergie par habitant, croissent, nous atteignons, à certains égards, les limites planétaires. Actuellement, personne ne doute que le changement climatique qui en résulte est réel et sérieux. Nous observons aussi d'autres problèmes environnementaux dus à l'activité humaine, tels que la surpêche ou la perte de biodiversité. En considérant ces phénomènes, on a l'impression qu' en tant qu'individus ainsi qu' en tant que sociétés , nous ne sommes pas capables de prendre soin de l'environnement dans lequel nous vivons.

Nous pouvons constater de grandes différences dans le comportement environnemental des personnes, ainsi que des pays, au travers des leurs politiques publiques mises en place. Bien que les raisons de ces différences aient été étudiées dans la littérature, certaines questions restent à explorer et ce sont elles qui ont poussé mon étude. Les questions auxquelles je tente de répondre structurent les trois chapitres de cette thèse.

D'abord, je tente de donner une nouvelle explication à la raison pour laquelle des pays similaires du point de vue du développement économique se comportent de manière différente par rapport aux polluants globaux, tels que les émissions de CO_2. Pourquoi des pays comme la Suède ou la Norvège (et les pays européens en général) se soucient davantage de leurs émissions de CO_2 tandis que d'autres pays comme les États-Unis ou le Canada continuent sur la même trajectoire d'émissions ? Pour répondre à cette question, je construis un modèle simple basé sur l'hétérogénéité de la population par rapport à ce que je appelle un 'trait morale' et, je montre que la société peut aboutir à une situation d'équilibre avec un comportement soit vert soit gris.

Dans le deuxième chapitre, j'opère un changement de champ en me situant à l'échelle mondiale. L'histoire nous confronte à une série d'échecs concernant l'obtention d'un traité international qui limite efficacement les émissions de CO_2. La littérature est riche à ce sujet et prédit un tel résultat : le problème auquel nous sommes confrontés est connu sous le nom de 'dilemme du prisonnier'. les pays préfèreraient collectivement de ne pas émettre, mais, en privé, chaque pays choisit de contaminer. Le résultat est que tous émettent. L'Eu-

rope est une exception puisqu'elle s'est imposée des restrictions sur ses émissions après l'échec du traité de Kyoto. Je propose dans le premier chapitre une explication des raisons pour lesquelles un pays peut arriver à adopter des politiques favorables à l'environnement. La question qui suit est donc : est-il possible pour l'Europe, ou d'autres pays verts, d'inciter les pays gris à abaisser efficacement leurs émissions ? Le deuxième chapitre explore cette question via la mise en place de deux instruments : un transfert technologique et une taxe aux frontières (border tax). Je montre que d'une situation comme celle-ci, il est possible d'induire une grande coalition qui abaisse efficacement les émissions. Le résultat montre également qu'il n'y a pas de solution unique quant au recours à la taxe, au transfert, ou aux deux en même temps, cela dépend des conditions particulières des pays (des paramètres du modèle).

Dans le troisième et dernier chapitre, je reviens au niveau national et essaie de comprendre comment les sociétés sont devenues ce qu'elles sont quant à leur comportement environnemental. J'utilise pour cela le concept de 'trait kantien' et observe comment il est distribué dans la société. Dans le premier chapitre, je suppose que cette distribution est donnée. Or, si nous pouvons changer la société vers une distribution plus 'kantienne', nous aurions une société qui se comportent de façon plus verte. Cela peut non seulement influencer cette société, mais aussi, comme le montre le chapitre deux, déclencher un revirement dans le comportement global. Pour répondre à cette question, je me sers d'un modèle connu de transfert intergénérationnel de trait. Ces modèles cherchent généralement à anticiper la composition d'une société dans le long terme en fonction des 'désirs' des parents de transférer certains 'traits' à leurs descendants. Je renverse la question et essaie de découvrir quels sont les 'désirs' des parents, à partir de l'observation d'une société donnée, avec sa distribution de qualité. Une meilleure compréhension du processus qui fait qu'une société est plus ou moins kantienne (par exemple) peut nous permettre de modifier le mécanisme qui façonne nos sociétés, et avec cela, d'essayer de les rendre plus vertes.

Je décris ensuite de manière détaillée pour chaque chapitre, l'obtention des résultats et les extensions possibles à réaliser.

En poussant vers des sociétés vertes : la motivation morale et le comportement vert

Ce chapitre vise à fournir une explication alternative des différences de comportement entre pays à l'égard de l'environnement et de la contribution à la pollution mondiale, bien que ces pays puissent être très similaires d'un point de vue économique et de développement. Pour explorer cette idée, j' utilise un modèle micro-fondé simple dans lequel les

individus retirent l'utilité de leur bien-être ainsi que d'une dimension morale. L'utilité de cette dernière notion vient de l'idée que les individus tirent satisfaction de faire 'la bonne chose' (au moins dans une certaine mesure) - ou, selon Emmanuel Kant, de se comporter selon le principe impératif. Être ou agir vert pourrait tomber dans la catégorie de ces principes impératives. Avec l'utilisation de ces concepts, en plus d'un cadre politique simple, je montre que deux sociétés équivalentes (i.e. les sociétés ayant le même revenu, le même système politique, etc.) peuvent atteindre deux équilibres de comportement environnemental différents. J'identifie aussi les moyens de pousser une société d'un équilibre vers un autre. Bien que je ne prétende pas que cette explication soit la seule raison pour laquelle les pays se comportent différemment, ce modèle offre une explication très simple pour laquelle cela pourrait se produire.

Différentes théories ont été développées pour expliquer, dans une certaine mesure, la dissemblance des comportements verts entre pays similaires. Un ensemble de théories tourne autour du niveau de développement d'un pays. Par exemple, la courbe environnementale de Kuznets (EKC) relie le comportement environnemental d'un pays à son niveau de revenu.[1] L'inconvénient d'une telle approche est qu'elle explique les différences concernant la pollution locale, mais qu'elle ne peut pas expliquer les différences de comportement concernant les polluants globaux (par exemple, les émissions CO_2 par habitant) entre pays ayant des niveaux de revenu similaires comme les États Unis, le Canada, l'Australie, et leurs homologues européens la Suède, la Norvège, et la Finlande.

Un autre pan important de la littérature explique le comportement vert fondé sur la motivation morale. Cela est dû au simple fait que les actions vertes ne peuvent être expliquées par une théorie de l'individu purement homo oeconomicus. Si les individus étaient de purs homo-oeconomicus, il n'y aurait pas d'incitation pour eux à contribuer au maintien d'un bon environnement, puisque le gain de cette action serait négligeable. Un point de départ dans la littérature est l'idée de 'warm-glow giving' (Andreoni (1990)), dans lequel le seul acte de donner fournit de l'utilité à l'agent. Dans une ligne similaire de pensée, Nyborg et al. (2006) présentent un modèle dans lequel les sociétés se polarisent en raison des effets de la pression des pairs (peer pressure). En d'autres termes, une société peut être complètement verte (quand tout le monde protège l'environnement) ou complètement grise (quand tout le monde choisit de polluer). Même si un tel résultat (extrême) n'a pas été observé dans la réalité, cette pression sociale existe.

[1] Le EKC est semblable à une courbe de Kuznets traditionnelle. Lorsqu' un pays devient plus riche et plus développé, il commence à polluer davantage (mesuré sur une base par habitant). Après avoir atteint un certain niveau de développement et de richesse, la société commence à accorder plus d'importance à l'environnement et commence à polluer moins. Ainsi, on observe une forme de U inversé pour la relation entre la pollution et le revenu par habitant. Bien que la EKC ait été utilisée pour analyser la pollution locale, elle pourrait également être utilisée pour comparer les pays concernant les polluants mondiaux.

Je propose une approche différente de la manière suivante : les individus sont diverses et se comportent, jusqu'à un certain degré, selon un motif proche de l'impératif kantien. Autrement dit, les individus essaient de faire 'la bonne chose', même si elle va à l'encontre de leurs intérêts privées (ou 'égoïstes'). Cette idée kantienne a été appliquée à l'économie dans Brekke et al. (2003), Brekke and Nyborg (2008), Nyborg (2011), Wirl (2011) et Laffont (1975). Les trois premiers articles utilisent l'approche kantienne afin de définir un idéal. A mesure que son comportement tend vers cet idéal, son image de soi grandit et l'individu gagne de l'utilité. De même façon, Wirl (2011) parle de joueurs verts qui trouvent l'optimum social suivant une approche kantienne. Dans le papier de Laffont (1975), les agents prennent en compte ce qui serait ou devrait être leur comportement en supposant que tout le monde va faire comme ils le font. Il trouve un équilibre de marché et montre que si les agents se comportent d'une manière kantienne, ils intériorisent leurs externalités et le résultat est optimal.

Dans mon modèle, je me réfère aux individus kantiens d'une manière similaire à Laffont. Ainsi, être "kantien" signifie que l'individu prend en considération ce qu'il "devrait" faire (la bonne chose), qui pourrait différer de ce qu'il ferait s'il agissait d'une manière purement "égoïste" (selon les principes de l'homo-oeconomicus). En ce sens, l'agent doit savoir ou apprendre ce qu' est la *"bonne chose"*. Il faut préciser que ce n'est pas exactement ce que Kant a voulu dire dans son travail (Kant et al. (2002)) ; Kant parle d' obligations (déontologie), pas de responsabilités. En ce sens, être kantien (dans la suite de ce document) signifie être une personne moralement responsable. Cela signifie faire une hypothèse sur les conséquences de nos actions (d'où le mot 'responsabilité') à l'aide d'une manière kantienne de penser – 'En supposant que tout le monde se comporte comme je le fais'.

En général, le choix d'être vert est un choix coûteux. Si il n'y avait pas de coût, il ne serait pas logique de pas le faire. Par conséquent, l'agent est confronté à un compromis entre faire la bonne chose et être affecté négativement par le coût direct de cette action. Ce coût est celui encouru lorsque l'on compare l'action verte avec celle qui serait faite de façon purement homo oeconomicus. Pour être en mesure d'évaluer efficacement la situation, l'agent doit savoir (ou estimer) 'son obligation morale' et son coût. Il pèse alors les options et décide comment se comporter. Comme nous pouvons le voir, ceci n'est pas une règle générale (et donc pas précisément kantienne), car une personne spécifique pourrait être influencé vers prendre part à un comportement vert ou gris en fonction du coût et de l'implication morale perçue de son choix. Par exemple, les implications de la contamination de l'environnement quand ce dernier est déjà très pollué sont tout à fait différentes de celles lorsque l'environnement est propre. Il est clair, cependant, que des personnes différentes pèsent le coût de leur action verte et de leur responsabilité morale de diverses

manières. Pour tenir compte de cela, je vais supposer que la société est composée d'un continuum de personnes, allant de l'individu purement homo-oeconomicus à celui purement kantien.

Comme nous pouvons le voir, ceci est un processus évolutif. La connaissance et les motivations morales des individus dépendent de la perception de la qualité de l'environnement (ou de la pollution). Les modifications de l'état de l'environnement et/ou des corrections de ses perceptions peuvent faire varier la sensibilisation. Dans le même temps, les décisions publiques sont prises par les gouvernements. Même si on suppose que les gouvernements tentent généralement d'appliquer les meilleures politiques, le cas de la pollution mondiale est un jeu de dilemme du prisonnier sur la scène internationale. Dans ce cas, il est 'optimale' pour le pays de polluer : c'est est la stratégie dominante du jeu. En ce sens, les gouvernements ne sont pas susceptibles de mettre en œuvre des politiques vertes (concernant les polluants mondiaux). Cependant, un gouvernement peut être disposé à assumer le coût de la mise en œuvre d'une politique verte s'il a le soutien de sa population, ce que la Suède a fait dans les années 1990.[(2)] En conséquence, j'introduis dans le modèle un mécanisme politique simple. Le gouvernement au pouvoir mettra une politique verte (ou plus verte) en place si les gens le demandent. Par conséquent, je suppose que si les individus se comportent de manière verte, leur gouvernement est également susceptible d'avoir une politique verte en place. Ceci concorde avec ce qui a été observé en Allemagne par Comin and Rode (2013), qui ont constaté que lorsque les gens se sont comportés de façon plus écologique, les partis verts ont reçu plus de votes lors des élections. Leiserowitz (2006) estime également que les individus 'egalitarian-value based' (similaires à ceux décrits comme incarnant le moral kantien) supportent les politiques vertes.

Enfin, je construis un modèle qui prend en considération les concepts mentionnés jusqu'ici. En combinant ces éléments – à savoir, les différents types de personnes à l'égard de leur comportement environnemental (leur 'structure' kantienne), le degré de conscience induite par l'environnement, et un système politique simple – je trouve que deux sociétés équivalentes peuvent atteindre deux équilibres environnementaux différents, même si elles partagent les mêmes caractéristiques structurelles (par rapport au revenu, au système politique, etc.). J' identifie également la possibilité de passer d'un équilibre à l'autre. En d'autres termes, il y a un point de basculement à partir duquel une société qui ne se comporte pas d'une manière respectueuse de l'environnement, peut être induite vers un comportement vert. Le mécanisme à l'œuvre repose sur l'idée qu' avoir un environne-

[(2)] La Suède a été le premier pays à introduire une taxe carbone en 1991, dont le taux a augmenté à travers le temps. Actuellement, ils visent à augmenter de nouveau le taux, afin d' atteindre la vision 2050 de zéro émissions nettes de GES' (gaz à effet de serre) de leur feuille de route du climat. (Energy Policies of IEA Countries, Sweden, 2013 review).

ment plus pollué rend les gens plus conscients, déclenchant le comportement vert. Il est possible que cela ne suffise pas à changer la praxis politique en vert. Toutefois, il pourrait aussi y avoir une situation dans laquelle les niveaux de sensibilisation et le comportement sont tels qu'un gouvernement vert (ou plus vert) est élu, déclenchant un processus de basculement du système. Ce cadre pourrait également expliquer pourquoi les pays avec des niveaux de revenu faibles se soucient de l'environnement parfois plus que les pays développés, comme indiqué dans Lee and Markowitz (2013).

En utilisant le résultat d'équilibres multiples, je montre également qu'il est possible de passer d'une trajectoire de "gris à vert". Fournir de l'information aux gens, par exemple, peut les sensibiliser à l'environnement et augmenter les chances de changement social. Ce résultat renforce les conclusions de Corbett and Durfee (2004) et Dunwoody (2007), qui montrent que les médias de masse ont une grande influence sur les préoccupations sociales, ce qui peut entraîner des changements dans le comportement environnemental d'un pays et, éventuellement, dans son cadre juridique.

Est-ce que cette idée de coup de coude est un concept qui pourrait être appliqué pour expliquer des comportements différents entre l'Europe et les Etats-Unis, à l'égard de questions environnementales telles que les émissions de dioxyde de carbone ? Ce pourrait être le cas de l'histoire des pluies acides de l'Europe qui a changé les perceptions et la sensibilisation sur la pollution transnationale européenne. Lorsque les pays européens se sont retrouvés confrontés à ces pluies acides causées par les émissions de leurs voisins, ils ont pris conscience du problème et ont établi des accords internationaux, spécifiques pour ces pays, sur l'environnement. Ca n'a pas été le cas en Amérique du Nord. Le problème des pluies acides a été essentiellement résolu par des actions à l'échelle nationale du gouvernement américain. Cela pourrait être un point de bifurcation à partir d'un chemin gris européen qui pourrait expliquer, dans une certaine mesure, la différence de comportement entre l'Amérique du Nord et l'Europe.

Le modèle actuel ne cherche pas à donner une explication complète de ce phénomène, mais plutôt de donner quelques idées alternatives de ce qui pourrait se produire. L'ensemble du modèle suppose que l' attitude kantienne de chaque personne est fixée. En d'autres termes, je suppose que les gens sont nés avec ce trait et qu'il n'y a aucune chance de changer au fil du temps. Évidemment, ceci est une hypothèse extrême. Cependant, l'idée ici est de créer un modèle qui peut expliquer les changements rapides dans le comportement vert observé au cours des dernières années.

Pourtant, nous pourrions introduire un second processus évolutif (plus lent) du trait kantien (plutôt que d'utiliser directement les changements de comportement). Ce proces-

sus pourrait être compris en considérant la façon dont l'éducation, par exemple, change les attitudes au fil du temps. Un tel changement serait du type décrit et modélisé dans Bisin and Verdier (2000). Les nouvelles générations abordent les questions environnementales d'une manière différente que les générations plus âgées, surtout quand ils reçoivent un enseignement sur l'environnement (et sur les complications de la dégradation de l'environnement) dans l'enfance. Un éventuel résultat empirique de cet effet est étudié par Hersch et Hersch and Viscusi (2005). Les auteurs montrent que les jeunes générations sont plus écologiques que les générations plus âgées.

Il pourrait également être intéressant d'explorer la relation entre les préoccupations des gens et leurs actions. Dans certains cas, il semble que le public soit 'préoccupé mais impassible', comme le montre Oppenheimer and Todorov (2006). Il semble que cela puisse être lié aux valeurs des gens (Leiserowitz (2006)). Cette relation pourrait être étudiée en utilisant un modèle avec une évolution des mentalités (ou traits).

Une autre extension possible concerne la relation entre le changement de la conscience sociale (et / ou de comportement) et un modèle plus complexe du système politique. Il serait enfin intéressant de vérifier la pertinence du modèle avec quelques informations empiriques. Néanmoins, certaines variables du modèle, telles la préoccupation des personnes et leur 'attitude kantienne', sont difficiles à mesurer correctement. Même ainsi, ce serait une belle aventure à poursuivre.

Pousser le basculement dans les accords environnementaux internationaux

Ce chapitre met l'accent sur l'idée que la création d'un accord international de l'environnement (AIE) qui rassemble de nombreux pays peut être une tâche colossale à accomplir. Barrett (1994), Barrett (2003), Rubio and Ulph (2006) et Eichner and Pethig (2013), entre autres, ont montré que le nombre de signataires de AIE auto-exécutoires ne dépasse pas trois ou quatre ; ou si ils le font, leurs émissions sont presque les mêmes que dans le Business as Usual (BAU). Les récents échecs pour parvenir à un accord concernant le changement climatique sont un exemple clair à cet égard. Les solutions analysées permettant d'arriver à un AIE couronné de succès impliquent de commencer par une coalition composée d'un petit sous-ensemble de pays, puis d'en intégrer plusieurs autres. Une première possibilité explorée par Heal and Kunreuther (2011) suggère d'agrandir une petite coalition dans un processus en cascade. Ils supposent un effet de renforcement positif parmi les signataires, qui induit l'adhésion de plusieurs autres pays. D'une manière similaire, Barrett (2006) a montré que si il existe une technologie verte qui a des rendements croissants à l'adoption, une coalition basculante émerge, que les autres pays rejoignent

pour finalement atteindre la grande coalition. Une deuxième ligne de pensée est analysée par Carraro and Siniscalco (1993), Hoel and Schneider (1997), and Barrett (2001), où les transferts sont utilisés afin d'inciter les pays à adhérer à l'AIE. Le premier travail présente un problème d'engagement qui soit empêche la formation de la grande coalition, soit, si celle-ci se forme, est composée de pays se comportant toujours comme dans le BAU. Le problème d'engagement vient du fait que le cadre de l'AIE ressemble à un jeu dit 'de la poule mouillée' (chicken game) : les signataires préfèrent cette situation plutôt que de ne pas arriver à un accord, mais ils préfèrent également s'abstenir et que d'autres signent. Par conséquent, quand les transferts sont effectués afin d'élargir la coalition, les signataires originaux préfèrent quitter. Hoel & Schneider (1997) ont trouvé le même effet, où l'engagement est acquis par la conformité. Barrett (2001) trouve un résultat similaire à Carraro & Siniscalco (1993) avec des pays asymétriques. Si l'asymétrie est faible, le problème d'engagement persiste. Si il y a une forte asymétrie, les transferts peuvent soutenir un résultat supérieur à celui de l'AIE sans paiements secondaires.

Avec toutes ces barrières, j'explore une nouvelle tactique. Je montre à l'aide d'une taxe à la frontière, d'un transfert technologique, ou les deux, qu'on peut induire et soutenir efficacement la grande et significative coalition.[3] Je trouve les conditions pour parvenir à chaque situation optimale. Il n'y a pas de recette générale, mais au moins un instrument est toujours mieux que rien. Aussi je montre que si les transferts font partie de la solution, ils doivent être faits pour les pays les moins verts. Ce choix améliore les chances de succès, tout en minimisant le montant des transferts. Ces résultats pourraient être importants face aux récents échecs pour parvenir à un AIE qui aborde efficacement le changement climatique. Il est possible que les raisonnements présentés dans ce travail puissent donner des solutions potentielles pour les négociations en cours.

Afin de développer cette idée, je construis un modèle qui est une version élargie de Barrett (1997). Dans ce travail, le monde est composé de pays identiques, contenant chacun une entreprise qui produit un bien homogène qui est échangé, et aussi de la pollution comme un produit secondaire. Dans le jeu de trois étapes de Barrett, chaque pays décide d'adhérer à un AIE ou non (one shot). Dans la deuxième étape, la coalition formée suit une logique de groupe ('group rational')groupe rationnel et joue par conséquent, avec les non-signataires qui agissent de manière indépendante. Enfin, les entreprises maximisent leurs profits en choisissant leurs output segmentés à la Cournot et le marché s'équilibre.

A partir de ce cadre, je commence par ajouter des asymétries entre les pays en termes de dommages causés par les émissions et les coûts d'abattement. Ces asymétries sont em-

[3]significative dans le sens que les pays font un effort pour baisser ces émissions et pas simplement agir comme dans le BAU.

piriquement motivées. Les sociétés présentent un comportement vert, même si la rationalité directe ne le laisse pas présager . Cela pourrait être lié à des raisons morales plutôt qu'économiques. Comme je montre dans le premier chapitre, des pays structurellement similaires (en termes de niveau de développement et le système politique) peuvent finir par se comporter différemment par rapport à l'environnement global. Cela pourrait être pour des raisons historiques, mais le fait est que certains pays sont devenus au courant du problème de l'environnement et ont commencé à agir en conséquence. Ainsi, les différences entre les pays sont leur sensibilité à la pollution (inquiétude environnementale) et le type de technologie d'abattement qu'ils ont. En ce qui concerne le premier point, une interprétation équivalente pourrait être la façon dont les pays sont affectés par la pollution. Alors, des dommages marginaux plus élevés provenant de l'émission seront traités comme synonyme d'être plus écologique. Du côté de l'abattement, les pays peuvent avoir une technologie mauvaise (coût marginal d'abattement élevé) ou bonne (faible coût marginal). Inversement, on peut penser que les pays historiquement verts ont investi dans les technologies propres (pour produire de l'électricité, par exemple), ce qui est équivalent à avoir une technologie avec un coût d'abattement pas cher. Il ressemble à l'Europe, qui a suivi cette voie, en investissant dans les technologies vertes et en s'imposant des contraintes sur les émissions. Suite à cela, je suppose qu'il y a deux groupes de pays : les pays riches, avec une inquiétude environnementale élevée et une technologie d'abattement à faible coût ; et les pays pauvres (aussi appelés 'outsiders') avec une inquiétude environnementale plus faible et des technologies d'abattement cher. Aussi, je suppose que pour les pays riches, il est rentable de former une coalition, ce qui n'est pas le cas pour les pays pauvres.

Deuxièmement, je rajoute la possibilité de transferts entre les pays, ce qui crée la possibilité d'agrandir la coalition existante, d'une manière similaire à Carraro and Siniscalco (1993). Cependant, je considère que les transferts sont des transferts de technologie (et non monétaires, comme font Carraro et Siniscalco), ce qui change le jeu. La dernière addition au modèle de base est que je permets à la coalition d'imposer une taxe à la frontière sur les non-signataires. Comme prévu, cette derrière idée est de dissuader le free-riding (et éventuellement induire l'adhésion), de façon similaire à l'interdiction du commerce dans le travail de Barrett (1997).

Avec cette configuration, je veux étudier si une petite coalition initiale de pays verts peut induire la formation de la grande coalition. L'idée derrière cela est la même que le mécanisme mis en œuvre avec succès dans le protocole de Montréal et ses modifications ultérieures (pour une explication détaillée, voir le livre de Barrett (2003)). Dans ce cas, un seul pays, les États-Unis, a été unilatéralement prêt à réduire les émissions et la consommation de substances appauvrissant l'ozone. Ils étaient bien conscients que les fuites pro-

venant du commerce pourrait réduire les résultats de leurs efforts, mais savaient que si les autres grandes économies suivraient, les gains (à la fois national et mondial) seraient beaucoup plus. Par conséquent, ils étaient prêts à non seulement interdire unilatéralement l'utilisation et la production de CFC, mais aussi d'interdire le commerce de ces substances. Cette deuxième composante, en plus du fait qu'ils étaient aussi prêts à aider les pays en développement pour passer à des substances nouvelles et propres, ont poussé d'autres pays à joindre à ce qui allait devenir l'une des plus réussies dans l'histoire de l'AIE.

Évidemment, il serait attrayant de faire la même chose avec l'effet de serre (GES). Cependant, alors que la lutte contre la production et le commerce des CFC est une tâche simple, il est tout à fait impossible pour les émissions de CO_2, car elles sont intégrées dans presque tous les produits que nous échangeons. Bien sûr, interdir complètement le commerce semble hors de question. Les 'bâtons' utilisés à Montréal ne sont pas crédibles dans un AIE concernant le CO_2. Pour éviter cet obstacle, nous pouvons utiliser une taxe à la frontière, imposée sur les marchandises en provenance de pays non-signataires, afin de décourager le free-riding et d'induire l'adhésion. Ce sera le cas dans le présent document, reconnaissant que cela pourrait déboucher sur une guerre commerciale. A cet égard, je suppose que les membres de la coalition peuvent imposer une taxe à la frontière et les non-signataires ne pas riposter. J'abandonne ensuite cette hypothèse et j'analyse si cela est crédible, et pour quels cas il détient effectivement. En outre, dans une étude récente de Nordhaus (2015), qui utilise un modèle DICE avec 15 régions spécifiques, l'auteur suppose également qu'il est possible d'imposer une taxe à la frontière sans représailles, et montre que cette taxe induit les pays à joindre la coalition. Dans mon travail, je résous pour quelles tailles de coalition et pour quelles conditions de paramètres, les non-membres ripostent. Je montre que pour les coalitions de petites et moyennes taille, les représailles ne valent pas la peine ; et que pour les grandes coalitions il dépend des conditions spécifiques sur les possibles non-membres. Je trouve aussi que la taxe imposée par les membres a une limite supérieure, afin d'induire la grande coalition. Prenant cela en considération, cela pourrait être le cas où une grande coalition, mais pas la grande coalition, devienne une solution stable.

Avec tout cela en tête, j'analyse si une taxe à la frontière imposée par les signataires, un transfert technologique, ou les deux, peut induire la grande coalition. Alors, j'étudie la stabilité de la grande coalition pour différents scénarios de paramètres. J'analyse pour quels cas soit l'impôt, soit le transfert, ou les deux ensemble, peut soutenir la grande coalition. Cette solution dépend, entre autres, des dommages environnementaux (ou inquiétudes) des pays. Ce résultat concorde avec le document susmentionné, depuis Nordhaus trouve aussi des effets mixés. Aussi, je suppose que les deux instruments sont nécessaires

et je trouve le groupe de destinataires optimal, qui est le groupe avec les pays les moins verts (dans les 'outsiders'). Je donne un exemple de ce cas, qui donne l'intuition de la façon dont les deux instruments fonctionnent ensemble.

Finalement, j'étudie le cas où nous pouvons rejeter la clause de participation minimale (CPM).[4] Ces clauses sont des outils juridiques qui aident les AIE à atteindre l'équilibre souhaité, par exemple, dans le cas de Montréal. Mais dans ce cas particulier, la CPM n'était pas trop élevé. Concernant le problème de CO_2, bien qu'il y ait un consensus sur les dommages résultants du réchauffement climatique, il y a des différences d'opinion entre les pays concernant la façon dont ces dommages pourront spécifiquement les toucher. En outre, le niveau de dommage est incertain et, éventuellement, comparable au coût de passage à la technologie verte nécessaire. L'incertitude sur les gains et les coûts d'un tel accord fait que la décision politique devient plus difficile, et met la CPM à un niveau qui pourrait ne pas être politiquement atteignable. Afin de comprendre ce qui devrait être mis en œuvre pour réduire la CPM nécessaire, et voir les interactions entre les deux instruments utilisés, je conduis un exercice numérique et vérifie sous quelles conditions la grande coalition est atteignable.

Une extension possible pourrait être fondée sur le papier de Nordhaus (2015). Son modèle est appliqué à un ensemble spécifique de paramètres (le monde est divisé en 15 régions) et le résultat général de ce travail pourrait aider à étudier la sensibilité de ses résultats à des changements dans cette configuration. D'autre part, il serait intéressant d'étudier la faisabilité des transferts de technologie, combinée ou non avec une taxe à la frontière, dans un modèle comme cela.

Enfin, une autre option peut être suggérée comme extension de mon travail. On pourrait analyser les implications stratégiques d'être un destinataire ou non. Cela vient du fait que les non-bénéficiaires ne reçoivent aucune incitation (car ils y adhèrent à cause de la pression de commerce). Par conséquent, cela pourrait les inciter à adhérer à l'accord dans un stade plus prématuré, et donc obtenir le transfert. Cette interaction stratégique peut aussi être prévue par les promoteurs de l'AIE et par le reste des joueurs, ce qui élargit le cadre de jeu et, éventuellement, peut changer ou réaffirmer le résultat précédent.

[4] La clause de participation minimale déclare que le traité devient obligatoire lorsqu'une quantité de pays supérieure ou égale à ce minimum ont signé l'accord. Depuis les AIE sont confrontés à des équilibres multiples (généralement deux) et afin d'induire la solution optimale, les traités comprennent une CPM. L'objectif c'est d'assurer que les signataires ne souffrent pas quand peu de pays ont signé l'accord.

L'évolution d'un 'Trait Kantien' : déduction depuis le jeu du dictateur

La littérature sur la transmission des traits sociaaux (ou valeurs sociales) utilise les outils de dynamique des populations afin de modéliser comment la prochaine génération sera, selon l'état actuel de la répartition des traits et des 'forces' impliquées dans le processus évolutif. Par exemple, Bisin and Verdier (2000) supposent que les agents de la génération actuelle ont une empathie myope lorsqu'ils considèrent l'utilité de leur progéniture et tentent ainsi de transmettre leurs propres traits à leurs enfants.[(5)] Cette empathie myope signifie que les parents évaluent l'utilité de leurs enfants en fonction de la leur. Cela implique aussi que les parents évaluent les actions de leurs enfants en utilisant leurs propres fonctions d'utilité. Si les traits des enfants diffèrent de ceux de leurs parents, il en résulte donc une utilité des enfants inférieure du point de vue des parents. J'utilise ce point de départ et je fais deux modifications : je suppose qu'il existe une 'aversion myope' (représentée par la fonction $v(\cdot)$) de l'agent vers le trait de son enfant lorsque ce trait est différent du sien ; et je suppose que ce trait peut être modélisé avec une variable positionnée sur une ligne continue entre zéro et un. Cette hypothèse se distingue de la littérature, les auteurs étudiant habituellement un scénario spécifique avec un nombre fini de types d'agents, le plus souvent égale à deux ou trois.

D'autre part, le jeu de dictateur (JD) est l'une des expériences les plus simples et les plus répliquées dans le domaine de la théorie des jeux. En un mot, l'expérience consiste à recruter deux personnes et de donner une somme d'argent (ou d'une autre chose de valeur) à l'un d'entre eux, choisi au hasard. La personne qui a reçu l'argent est appelée le dictateur et il lui est demandé de partager une certaine fraction (ou pas) de cet somme avec la seconde personne, appelée le bénéficiaire. Après cela, l'argent est réparti selon cette décision. Par conséquent, le bénéficiaire n'a rien à dire dans ce jeu ; nous sommes seulement intéressés par la décision du dictateur. Selon la théorie homo-oeconomicus, le dictateur ne devrait jamais donner un centime, mais ces expériences montrent une quantité constante et considérable de personnes partageant une partie de l'argent, même à une proportion de 50/50 ou plus. Différentes explications de ce phénomène sont avancées. En supposant que le dictateur n'a aucune relation directe ou indirecte avec le bénéficiaire, qui est généralement le cas testé, ces explications pointent l'idée d'une norme sociale et / ou un trait de moralité, qui sont tous deux transmis entre les générations. Il y a d'autres expériences (jeux) qui pourraient être utilisés pour en déduire les traits de ce genre.

Suite à l'idée du premier chapitre concernant une 'morale kantienne', où les agents sont dotés d'un trait kantien qui se trouve dans un spectre continu, je peux tracer les réponses du JD à un niveau de trait kantien. Dans ce cadre, je définis une personne com-

[(5)] Ce type de transmission a également été utilisé dans Bisin and Verdier (2001), Hauk and Saez-Marti (2002) et Sáez-Martı and Zenou (2012).

plètement kantienne comme un agent qui maximise son utilité *en supposant* que tout le monde agit comme il le fait, en contraste avec l'agent homo oeconomicus qui maximise juste son utilité de manière égoïste. Il est important de noter que cela textbfn'est pas ce que Kant voulait dire dans son travail séminal. J'utilise le terme 'kantien' de la même manière que Laffont (1975). L'idée ici est que chaque personne veut se comporter d'une manière moralement responsable, au moins dans un certain degré. Par conséquent, si le degré est plein, je l'appelle cet agent d'une personne entièrement kantienne. Si le degré est nul, nous avons le cas homo-oeconomicus pur. Dans cette interprétation, comme dans Laffont (1975), la personne entièrement kantienne fait son choix sous l'hypothèse que tout le monde se comporte comme il le fait. Bien sûr, elle ne prévoit pas que cela se produise ; il est juste la façon empirique de trouver la 'bonne chose à faire'. Évidemment, ce n'est pas ce que le catégorique impératif est, mais j'emprunte le nom parce que la démarche est proche de l'idée de Kant. Les degrés entre les deux extrêmes reflètent des gens qui donnent un certain poids à la responsabilité kantienne (que je nomme α) et le reste à l'attitude homo-oeconomicus (pondérée par conséquent $(1-\alpha)$). Par conséquent, nous pouvons avoir des gens qui se comportent d'une manière intermédiaire, qui est capturé par ce trait kantien continue. Bien évidemment ce n'est pas exactement ce que Kant voulait dire dans son travail séminal, il tente de refléter ce que nous observons dans la vie réelle ; par exemple, en utilisant les résultats des expériences de JD, nous donne une grande variété de résultats. Un résultat intéressant des expériences de JD met en évidence une polarisation de la distribution, avec à peu près un tiers de la population agissant d'une manière purement homo-oeconomicus (qui ne donnent rien), un autre tiers partageant la moitié de la somme (ou même plus), que j'interprète comme complètement kantien, et le reste des individus répartis, presque uniformément, entre ces deux extrêmes.

Avec ces deux éléments, un trait kantien et un système 'évolutif' de traits, l'objectif de ce travail est de rationaliser l'évolution de la distribution du trait kantien qui pourrait expliquer les résultats des expériences de JD.[6] Pour ce faire, je vais d'abord développer un modèle continu de la dynamique des populations, en utilisant le modèle discret comme point de départ. Je suppose qu'il existe une fonction $v(\cdot)$ qui représente 'l'aversion', ou la perte en utilité du parent, d'avoir un enfant avec un trait différent du sien. L'input de la fonction est la différence de trait de l'enfant par rapport à la caractéristique de son parent. La fonction $v(\cdot)$ n'a pas besoin d'être symétrique en zéro. En fait, nous verrons que pour adapter les résultats des expériences de JD, elle ne sera pas symétrique. Avec cela, je trouve les conditions d'équilibre de la dynamique où la distribution de la population cesse d'évoluer. Je suppose ici que les résultats des expériences de JD sont en fait

[6]Le mot 'évolution' utilisé ici a ses origines dans des études en biologie, où les biologistes ont analysé comment évoluaient différentes populations animales en fonction de la présence d'autres espèces. Il ne fait pas référence à l'évolution génétique, mais, puisque les processus sont proches en substance, il a hérité le nom.

à l'équilibre, ou tout près de celui-ci. Avec ce cadre, je montre quelques résultats simples concernant les solutions possibles de fonctions $v(\cdot)$, qui évoluent dans une distribution d'équilibre donnée de la population. Cette répartition se situe dans le segment continu $[0, 1]$, et il peut y avoir des points de concentration élevée, qui sont représentés avec des deltas de Dirac. La seule condition pour la distribution est que son intégrale doit être égale à un, comme dans toute distribution.

Avant d'obtenir les résultats sur les conditions d'équilibre et sur certaines propriétés de $v(\cdot)$, j'essaie ensuite de trouver des fonctions $v(\cdot)$ qui peuvent évoluer dans une distribution de la population plus conforme aux résultats des expériences de JD. La résolution analytique de ce problème n'est pas possible pour ce cas asymétrique et je procède donc par simulations. Je développe un algorithme qui prend la distribution finale (résultante) et une estimation initiale de $v(\cdot)$ comme inputs. En utilisant un processus itératif, l'algorithme converge vers une fonction $v(\cdot)$ qui satisfait les conditions d'équilibre. Dans le cas spécifique d'un trait kantien et en utilisant des résultats des expériences de JD, je trouve que les individus homo-oeconomicus ont une aversion plus forte, ou une désutilité d'avoir un enfant avec un trait différent plus forte, que leurs homologues kantiens. Suite à l'origine de la volonté des parents qui explique la transmission de trait, on arrive à que l'empathie myopique est (au moins une) raison de vouloir nos enfants à être comme nous sommes. Ainsi, les individus homo-oeconomicus semblent être plus myopes que les individus kantiens, ce qui a du sens si on considère les définitions des attitudes kantiennes et homo-oeconomicus. Les personnes homo-oeconomicus ont tendance à être plus égoïstes tandis que les kantiens sont plus empathiques. En supposant que toutes ces hypothèses soient vraies, il s'avère ironique que les individus kantiens, en prenant davantage soin de leurs semblables, mettent plus en péril leur propre existence (évolutive).

Il serait intéressant de mieux comprendre les origines de la fonction $v(\cdot)$, responsable des forces de transmission qui façonnent notre société, au moins dans ce cadre. Une meilleure compréhension pourrait aider à mieux comprendre comment inciter les sociétés à devenir plus vertes. Une deuxième ligne de recherche possible est d'essayer de trouver une solution mathématique plus générale de la fonction $v(\cdot)$.

Ayant constaté différentes distributions de trait pour les différents pays (comme, par exemple, la distribution de trait kantien), il pourrait également être intéressant d'examiner si ces différences sont liées à d'autres résultats dans ces pays, tels que, par exemple, leurs caractéristiques concernant la conservation de l'environnement, des résultats d'autres jeux, ou le comportement vert réel. Si l'idée précédente se révèle être vraie, alors on pourrait étudier comment modifier la fonction $v(\cdot)$ ou le coût de la transmission d'un trait kantien, ou quelque chose le long de ces lignes afin d'induire et de soutenir le change-

ment vers des sociétés vertes.

Le trait kantien discuté ci-dessus peut être lié à la façon dont les gens sont généreux et comment cette générosité est transmis de génération en génération. D'autres traits sociaux pourraient être analysés de la même façon. Une extension plus assez directe pourrait donc être entreprise en étudiant les motivations 'intrinsèques' pour d'autres types de traits.

Une dernière idée à explorer consisterait à relâcher l'hypothèse que la société est à l'équilibre. Nous devrions vérifier si les différentes cohortes se comportent (significativement) différemment les uns des autres. Si elles ne le font pas, les résultats actuels pourraient être vrais. Si elles le font, nous pourrions explorer comment les cohortes changent (vitesse et forme). Il pourrait ainsi être possible de dériver la fonction $v(\cdot)$ et éventuellement d'autres propriétés intéressantes.

Bibliographie

James Andreoni. Impure altruism and donations to public goods : A theory of warm-glow giving ? *Economic Journal*, 100(401) :464–77, June 1990.

Scott Barrett. Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46 :pp. 878–894, 1994. ISSN 00307653.

Scott Barrett. The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics*, 19(4) :345–361, November 1997.

Scott Barrett. International cooperation for sale. *European Economic Review*, 45(10) :1835–1850, December 2001.

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

Scott Barrett. Climate treaties and "breakthrough" technologies. *American Economic Review*, 96(2) :22–25, 2006.

Alberto Bisin and Thierry Verdier. A model of cultural transmission, voting and political ideology. *European Journal of Political Economy*, 16(1) :5–29, March 2000.

Alberto Bisin and Thierry Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2) :298 – 319, 2001. ISSN 0022-0531.

Kjell Arne Brekke and Karine Nyborg. Attracting responsible employees : Green production as labor market screening. *Resource and Energy Economics*, 30(4) :509–526, December 2008.

Kjell Arne Brekke, Snorre Kverndokk, and Karine Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9-10) :1967–1983, September 2003.

Carlo Carraro and Domenico Siniscalco. Strategies for the international protection of the environment. *Journal of Public Economics*, 52(3) :309–328, October 1993.

Diego Comin and Johannes Rode. From green users to green voters. Working Paper 19219, National Bureau of Economic Research, July 2013.

Julia B Corbett and Jessica L Durfee. Testing public (un) certainty of science media representations of global warming. *Science Communication*, 26(2) :129–151, 2004.

Sharon Dunwoody. The challenge of trying to make a difference using media messages. *Creating a climate for change*, pages 89–104, 2007.

Thomas Eichner and Rüdiger Pethig. Self-enforcing environmental agreements and international trade. *Journal of Public Economics*, 102(0) :37 – 50, 2013. ISSN 0047-2727.

Esther Hauk and Maria Saez-Marti. On the cultural transmission of corruption. *Journal of Economic Theory*, 107(2) :311 – 335, 2002. ISSN 0022-0531.

Geoffrey Heal and Howard Kunreuther. Tipping climate negotiations. Technical report, National Bureau of Economic Research, 2011.

Joni Hersch and W. Kip Viscusi. The generational divide in support for environmental policies : European evidence. NBER Working Papers 11859, National Bureau of Economic Research, Inc, December 2005.

Michael Hoel and Kerstin Schneider. Incentives to participate in an international environmental agreement. *Environmental and Resource Economics*, 9(2) :153–170, 1997. ISSN 0924-6460.

I. Kant, A.W. Wood, and J.J.B. Schneewind. *Groundwork for the Metaphysics of Morals*. Rethinking the Western Tradition. Yale University Press, 2002. ISBN 9780300094879.

Jean-Jacques Laffont. Macroeconomic constraints, economic efficiency and ethics : An introduction to kantian economics. *Economica*, 42(168) :430–37, November 1975.

Tien Ming Lee and Ezra Markowitz. Disparity in the predictors of public climate change awareness and risk perception worldwide. 2013.

Anthony Leiserowitz. Climate change risk perception and policy preferences : the role of affect, imagery, and values. *Climatic change*, 77(1-2) :45–72, 2006.

William Nordhaus. Climate clubs : Overcoming free-riding in international climate policy. *American Economic Review*, 105(4) :1339–70, 2015.

Karine Nyborg. I don't want to hear about it : Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization*, 79(3) :263–274, August 2011.

Karine Nyborg, Richard B. Howarth, and Kjell Arne Brekke. Green consumers and public policy : On socially contingent moral motivation. *Resource and Energy Economics*, 28(4) : 351–366, November 2006.

M. Oppenheimer and A. Todorov. Global warming : The psychology of long term risk. *Climatic Change*, 77(1-2) :1–6, 2006. ISSN 0165-0009.

SJ Rubio and A Ulph. Self-enforcing agreements and international trade in greenhouse emission rights. *Oxford Economic Papers*, 58 :233–263, 2006.

Maria Sáez-Martı and Yves Zenou. Cultural transmission and discrimination. *Journal of Urban Economics*, 72(2) :137–146, 2012.

Franz Wirl. Global warming with green and brown consumers. *Scandinavian Journal of Economics*, 113(4) :866–884, December 2011.

Chapter 1

Moving toward Greener Societies: Moral Motivation and Green Behaviour

1.1 Introduction

This paper aims to provide an alternative explanation for why countries behave differently with respect to the environment and contributions to global pollution, although they might be quite similar from an economic development point of view. To explore this idea, I use a simple micro-founded model in which individuals derive utility from their own well-being as well as from a moral standpoint. The utility of the latter concept comes from the idea that individuals derive satisfaction from doing 'the right thing' (at least to some degree) – or, according to Immanuel Kant, from behaving according to the imperative principle. Being or acting green could fall into the category of such imperative principles. Using these concepts in addition to a simple political framework, I show that two equivalent societies (i.e., societies with the same income, political system, etc.) can reach two different environmental behaviour equilibria. I also locate the means of nudging a society from one equilibrium to another. Although I do not claim that this explanation is the only reason for why countries behave differently, this model provides a very simple rationale for why this could happen.

Different theories have been developed to explain, to some extent, the dissimilarity of green behaviour among similar countries. One set of theories revolves around a country's level of development. For example, the Environmental Kuznets Curve (EKC) relates a country's environmental behaviour to its income level, revealing an inverted U-shaped relationship between the two factors.[(1)] The drawback of such an approach is that

[(1)]The EKC is similar to a traditional Kuznets curve. As a country gets richer and more developed, it begins to pollute more (as measured on a per-capita basis). After reaching a certain developmental level, society begins to designate more importance to the environment and therefore starts to pollute less as it

it explains dissimilarities concerning local pollution, but it cannot explain the dissimilar behaviour when it comes to global pollutants (for example, CO_2 emissions per capita) among countries with similar income levels like the USA, Canada, Australia, and their European counterparts Sweden, Norway, and Finland.

Another important strand of literature explains green behaviour based on moral motivation. This is due to the simple fact that green actions cannot be explained with a purely homo-oeconomicus theory. If individuals only abided by homo-oeconomicus principles, there would be no incentive for them to contribute to maintaining a good environment, since the gain for this action would be negligible. One starting point in the literature is the idea of 'warm-glow giving' (Andreoni (1990)), in which the sole act of giving provides utility to the agent. This notion has been further developed by Nyborg and Rege (2003) and Nyborg et al. (2006). The former study provides a comprehensive summary of different types of moral motivations (altruism models, social norm models, fairness models, models of commitment, and the cognitive evaluation theory), although it focuses on how these different types of moral motivations can crowd out private contributions. Nyborg et al. (2006) present a model in which societies become polarized due to the effects of peer pressure. In other words, a society can be completely green (when everyone protects the environment) or completely grey (when everyone chooses to pollute). While such an (extreme) outcome has not been observed in reality, it is clear that this social pressure exists.

I propose a different approach as follows: People are diverse and behave, up to some degree, according to something close to a Kantian imperative. That is, individuals try to do 'the right thing', even if it goes against their private (or 'selfish') homo-oeconomicus tendencies. This Kantian idea has been applied to economics in Brekke et al. (2003), Brekke and Nyborg (2008), Nyborg (2011), Wirl (2011), Roemer (2010), and Laffont (1975). The first three papers use the Kantian approach in order to define an ideal. In measuring his behaviour against this ideal, one gains utility from his self-image, which grows as his behaviour approaches the ideal. Similarly, Wirl (2011) talks about green players that find the social optimum following a Kantian approach. Roemer's paper (2010) uses the Kantian idea to create a rule in a Kantian manner. This is a game-theoretic approach which defines a Kantian equilibrium as one that no one wants to deviate from in the same *proportion*. In Laffont's paper (1975), agents take into account what would or should be their behaviour with the premise that everyone will do as they do. He then derives a market equilibrium and shows that if agents behave in a Kantian way, they internalize their externalities and the outcome is optimal.

becomes richer. Hence, we observe an inverted U-shape for the relationship between pollution and income per capita. Although the EKC was used to analyse local pollution, it could also be used to compare countries concerning global pollutants.

In my model, I refer to Kantian people in a way similar to Laffont. Here Kantian means that I consider what I should do (the right thing), which might differ from what I would do if I were to act in a purely homo-oeconomicus way. In this sense, the agent has to know or find out what *the right thing* is. This is not exactly what Kant intended in his work (Kant et al. (2002)); he described duties (deontology), not responsibilities. He talked about a *maxim* (universal law) instead of *actions*. However, it may be argued that one succeeds or fails at meeting their responsibilities as a consequence of his or her actions. Now, if agents thought that their contribution to pollution (or its prevention) was negligible, they would not say (in a direct way) that they were responsible for climate change. However, it is widely accepted that we are all responsible. In that sense, to be Kantian (in this paper) means being a morally responsible person. It means making an assumption about the consequences of our actions (hence the word 'responsibility') using a Kantian way of thinking – "assuming everyone behaves as I do". We often hear others make statements like: *"I don't pollute (or I do recycle, etc.) because if everyone did so, the effects would be terrible, and I don't want to contribute to that"*. This tells us that the person has assessed the situation, made a decision that the action is morally bad, and is taking the action of refraining from participating in it if at all possible. This practice of responsibility in moral systems may be explained by a variety of primary ethical principles, including the categorical imperative (Kant), utilitarianism, contractualism, cooperation, compassion, etc., as explained in Baumgärtner et al. (2014). In the present paper, I use the Kantian categorical imperative as a primary ethical principle. Supporting this idea, I can mention Bruvoll et al. (2000). They find that 88% of surveyed people in Norway, when asked about their motives for sorting waste, agreed or partly agreed with the following statement: *"I recycle partly because I think I should do what I want others to do"*. This is definitely a Kantian motivation.

In general, the choice to be green is a costly one; if it were not, of course, it would not make sense *not* to do so. Therefore, the agent is faced with a trade-off between doing the right thing and being negatively affected by the direct cost of this action. This cost is the one incurred when we compare the green action with the one that would be made in a purely homo-oeconomicus way. To be able to effectively assess the situation, the agent must know (or estimate) his moral 'obligation' and its cost. He then weighs the options and decides how to behave. As we can see, this is not a general rule (and therefore not precisely Kantian), since a specific person could be swayed toward taking part in green or grey behaviour depending on the cost and the *perceived* moral implication of his choice. For example, the implications of contaminating the environment when it is already quite polluted are quite different from those in a situation in which the environment is clean. It is clear, though, that different people weigh the cost of their green action and their moral responsibility in diverse ways. To account for this, I will assume that society is composed

of a continuum of people, ranging from the purely homo-oeconomicus individual to the purely Kantian one.

As we can see, this is an evolving process. Individuals' awareness and moral motivations depend on the perception of environmental quality (or pollution). Changes to the environmental state and/or corrections to one's perceptions can cause awareness to vary. At the same time, public decisions are made by governments, which generally try to apply the best policies. The problem arises from the fact that global pollution is usually a prisoner's dilemma game (in the international arena), and in this case, it is 'optimal' for the country to pollute: It is the dominant strategy of the game. In this sense, governments are not likely to implement green policies (concerning global pollutants). However, a government may be willing to incur the cost of implementing a green policy if it has the support of its constituency, as Sweden did in the late 1990s.[(2)] Consequently, I introduce a simple political mechanism into the model. The government in power will put a green (or greener) policy in place if the people demand it. Therefore, I assume that if individuals behave in a green way, their government is also likely to have a green policy in place. This is in line with what was observed in Germany by Comin and Rode (2013), who found that when people behaved in a greener way, green parties received more votes at elections. It is also in line with Leiserowitz (2006), who finds that *"egalitarian-value based"* individuals (similar to those described as embodying the Kantian morale) stand for green policies.

On the perception side, a poorer environmental state, which is expressed by Mother Nature through more frequent and severe climate events, triggers concern. This fact has been comprehensively studied by different surveys (as in Gallup and GlobeScan, among others: Gallup world poll [25], Globescan radar [26] and Extreme Weather and Climate Change in the American Mind April 2013 [31]) and verified using econometric techniques by Krosnick et al. (2006) and Zahran et al. (2006). Lee and Markowitz (2013) performed an overall analysis showing that individuals' environmental awareness and concern rises after major climate events, although typically only in the short term. As an illustrative example, after superstorm Sandy hit New York City President Obama said: "We must do more to combat climate change ... Now, it's true that no single event makes a trend. But the fact is, the 12 hottest years on record have all come in the last 15. Heat waves, droughts, wildfires and floods – all are now more frequent and more intense."[(3)] Although the president's making a statement does not necessarily guarantee that proper environmental behaviour will materialise immediately, it shows that awareness of cli-

[(2)]Sweden was the first country to introduce a carbon tax in 1991, and they have been increasing it as time goes by. Currently, they are aiming to increase it again, in order to "point out how to achieve the 2050 vision of zero net GHG emissions" in their Climate Roadmap. (Energy Policies of IEA Countries, Sweden, 2013 review).

[(3)]President Obama at the State of the Union Address. February 12th, 2013.

mate issues has reached a significant segment of the population, and that these issues have been acknowledged by those in power.

Finally, I build a model that takes into consideration the concepts mentioned so far. By combining these elements – namely, different types of people with regard to their environmental behaviour (their Kantian 'structure'), the degree of environmentally induced awareness, and a simple political system – I find that two equivalent societies can reach two different environmental behaviour equilibria even if they share the same structural characteristics (with respect to income, political system, etc.). I also identify the possibility of switching from one equilibrium to another. In other words, there is a tipping point at which a society that is not behaving in an environmentally-friendly manner can be swayed toward green behaviour. The mechanics are related to the idea that having a more polluted environment makes people more aware, triggering green behaviour. It may be the case that this is not enough to switch the political praxis to a green one, and we could be left with a grey society; however, there could also be a situation in which the awareness levels and behaviour are such that a green(er) government gets elected, starting a process to tip the system.

The logical question that arises is: How can a society be swayed from grey to green? To answer this, I analyse the influence of two factors: individuals' perception of pollution and the existing political system. A shock to the perception of pollution (such that individuals become more aware of or concerned with environmental issues) is an effective mechanism to induce tipping. Moreover, a political framework in which coalitions are more likely to exist eases the shift from a grey to green society and vice-versa. This is mainly due to the fact that a more 'continuous' political spectrum allows a society to shift toward a relatively greener government and, from there, to greener and greener governments in a cascading process. Applied at a government level, this cascading idea is similar to the one developed by Kuran (1991).[(4)] Continuing with the political angle, the literature has also addressed the determinants of green behaviour by comparing political systems.[(5)] Unfortunately, these results present neither a clear nor consistent view of how

[(4)]Kuran talks about the collapse of Eastern Europe's communist regimes. He divides the society into 10 types of people, ranging from those who are more in favour of a communist government, to those completely opposed to it. He shows that if some sort of threshold is crossed, protests can begin, which can encourage those initially less likely to go against the incumbent regime to join in protesting. This process can lead to a cascading effect, which can in turn trigger the collapse of the whole regime.

[(5)]Persson et al. (2000) used a theoretical model to show that presidential regimes should produce an under-provision of public goods (thus leading to a dirtier environment). On the other hand, Bernauer and Koubi (2004) found the opposite result. They use an econometric study to find evidence that presidential democracies provide more public goods than do parliamentarian democracies. More recently, Saha (2007) tested the previous hypotheses empirically. She finds that the electoral system has no effect on any of the environmental public good supply indicators and that the nature of the political regime has no significant impact either.

political systems might influence a society's green behaviour. Since the previous idea is closely related to this literature, it might explain, at least partially, the results found in this literature.

To finalise the model, and acknowledging the existence of another psychological ingredient, I add to the model *peer effects* or *social approval*. They can act as secondary motivators, as in the case of Nyborg et al. (2006). Incorporating this concept into the model reveals that, in fact, an 'ideological' peer effect[6] makes the transition from a grey to green society a more difficult task to achieve. This comes from the notion that if the society is primarily grey, the agent will have to bear his economic cost ***plus*** the new peer pressure cost in order to behave in a green way, thus making the shift harder to accomplish.

The chapter is structured as follows: First, Section 1.2 presents the model and its main features. Section 1.3 shows possible tipping points and demonstrates how the system can be nudged. Section 1.4 introduces the concept of social approval as a psychological driver of behaviour. Section 1.5 concludes and presents a brief discussion of the model.

1.2 The Model

The people

I assume that people care about their own utility and about social utility, as mentioned in the Introduction. The fact that individuals can have different ***attitudes*** regarding social well-being causes heterogeneity among agents. Therefore, we can write each agent's utility as:

$$U(\cdot) = (1 - \alpha)\, u_p + \alpha\, u_s \tag{1.1}$$

where u_p and u_s are the private and social utility respectively. The parameter α ($0 \leq \alpha \leq 1$) represents how homo-oeconomicus ($\alpha \to 0$) or Kantian ($\alpha \to 1$) the person is, hence his attitude. An attitude can be defined as an inherent trait formed from a combination of cultural background and education. It relates to and influences an agent's moral responsibility. For the purposes of this study, $\alpha = 0$ means that the agent does not care at all for the rest of the society, whereas $\alpha = 1$ means that an agent is the most morally responsible (Kantian). Moreover, the society is composed of a continuum of people, finite in number, each one matching a value of α in a biunivocal correspondence where α has some distribution f_α.[7]

[6] This means that green people prefer others to behave in a green way as well, and grey people prefer others to behave in a grey way. Of course, it might be the case that being grey is always considered 'bad', even for grey people, as in the case of smoking.

[7] As will be made clear in the following pages, the distribution of α will not change the main results, but assuming a uniform distribution will certainly ease the subsequent calculations and simulations. The

Both u_p and u_s are constructed in the same way. They have a consumption part $u(\cdot)$ and a damage part $d(p_t(\cdot))$:

$$u_i = u(\cdot) - d(p_t(\cdot)) \tag{1.2}$$

The first part, $u(\cdot)$, is the classic consumption utility with $u' > 0$ and $u'' < 0$. The damage term $d(p_t(\cdot))$ also has its classic properties of $d' \geq 0$, $d'' > 0$ and $d'(0) = 0$, where $p_t(\cdot)$ denotes the pollution level at time t. I return to these functions in the following pages.

Goods and Pollution

In this simple framework, each person can either buy green products (x) or grey products (y). From a consumption point of view, the products are perfect substitutes. However, there are two differences. The first is that the green one does not pollute, whereas the grey one does. The other difference is that the green product is more expensive than its grey counterpart.[8] I assign a normalized price of 1 to the grey good and a price of $(1+\rho)$ to the green one. Therefore, the value of ρ represents the extra amount (with respect to the whole original price) to be paid for a green product. Since the grey product pollutes, I also denote with γ the impact on the environment of the consumption (or more accurately, production) of this type of product. For simplicity, the agent will only choose one or the other, not a mix.[9] The agent's income is also normalized to 1.

Hence, the agent can be a 'grey' consumer, $(x, y) = (0, 1)$, or a 'green' consumer, $(x, y) = (\frac{1}{1+\rho}, 0)$. The pollution equation will be the standard one:

$$p_t = (1-\delta)p_{t-1} + \gamma \cdot y_t^s \tag{1.3}$$

where p_t is the pollution level at time t, δ is the natural decay of pollution level (due to natural absorption), and y_t^s is the *society's* mean grey consumption at time t, which is just the average of grey consumption by all agents. I will come back to this term in the following pages.

analysis will be performed using a uniform distribution. On the other hand, different distributions will simply change the place of the tipping point and the conditions needed to tip, as in Kuran (1991).

[8] If there were a green good that was cheaper than its grey counterpart, agents would automatically choose that good instead of the grey one for purely economic reasons. If that were the case, we could recalculate the pollution produced by a new representative grey good and return to the set-up presented here.

[9] This actually does not make any difference, except for the extra complexity. If the agent can choose a mix of green and grey products, all results *hold*. The behaviour functions and the resulting dynamics turn out to be the same. A more extended explanation and mathematical development can be obtained upon request.

Public Concern

I will call 'public concern' the part of the utility function that pertains to social welfare. Each agent is weighted by the parameter α, which captures the degree to which he or she is concerned about public well-being or the relative weight they assign to being morally responsible. In order to model the public concern and the impacts of this concern on the agent's behaviour, I use the Kantian morale previously described. More precisely, the agent will consider u_s as if everyone else were behaving as he or she is. This does not mean that the agent is actually expecting that everyone will behave exactly as he or she does. However, it allows us to mimic the decision process by modelling the utility of doing the right thing, which is in line with the Kantian idea. In other words, this part of the utility function is modelled as though the agent assumes that everyone is behaving as he or she does in order to make a decision about how to behave. Strictly speaking, this is not what Kant meant with his categorical imperative, as already discussed in the Introduction. His was not an heteronomous ethic. In this sense, the present formulation is not categorical, autonomous or independent of external influence; on the contrary, it depends on the environment's quality. Rather, my formulation is in line with the one used by Laffont (1975), which borrows the idea of choosing the good (ethical) rule when assuming that everybody behaves as oneself does hence the term 'Kantian morale'.

Regarding pollution, the level considered by the agent is the (estimated) pollution level p_t^e. This, in turn, depends on two factors: the perceived (past) pollution level, p_{t-1}^p, and the assumed emissions. At this point I assume that the agent has perfect information about the past pollution level, $p_{t-1}^p = p_{t-1}$. Recalling the pollution equation (1.3), we get the following relationship:

$$p_t^e(y^s) = (1-\delta)p_{t-1}^p + \gamma \cdot y^s \tag{1.4}$$

Now I return to the private and social utility functions. For u_p, the agent understands that he is atomistic with respect to the society, and hence he knows that his contamination is negligible with respect to the total emissions. This translates to a damage term $d(p_t(y))$ that does not vary with his individual decision y, but depends only on the society's behaviour y^s. Since I will compare the agent's two options, and because the previous term does not vary with the agent's decision, I drop it from the following equations (it cancels out).

On the other hand, u_s is the social utility taken into account by the agent when using a Kantian view. In other words, the agent considers that everyone behaves as he or she does, implying that society's emissions y^s will follow their choice y, as well as x^s with x. Putting all of the pieces together and rewriting the previous equations, we get:

$$u_p = u(x,y) \qquad \text{and} \qquad u_s = u(x,y) - d(p_t^e(y^s = y)) \tag{1.5}$$

Finally, we can plug these results into the agent's utility function 1.1, arriving at:[10]

$$U(y) = (1-\alpha)\cdot[u(x,y)] \; + \; \alpha\cdot[u(x,y) - d(p_t^e(y^s = y))]$$

$$U(y) = u(x,y) - \alpha\cdot d(p_t^e(y)) \tag{1.6}$$

$$U(y) = u(x,y) - \alpha\cdot d[(1-\delta)p_{t-1} + \gamma\cdot y] \tag{1.7}$$

This formulation is in harmony with the standard representation of green behaviour. Since a purely homo-oeconomicus approach cannot explain this type of behaviour, we must consider a moral motivation, as stated in the Introduction. People behave in a green way because they think it is the right thing to do, not because it is in their best economic interest to do so. But what is the right thing to do in a framework like this one? To tackle this question, I have used the ideas of Immanuel Kant. In his exploration of what was 'good' and 'bad', he devised the idea that a good action was one that could be tested as a maxim rule, one that everyone would follow, known as the categorical imperative (Kant et al. (2002)). If this rule makes the society better off, then it is a good rule to follow, meaning that following it makes us good people. In order to use this idea, for the present formulation, this dictum can be translated into: *Which general rule of action should I follow to maximize social welfare, as I perceive it, given that everyone acts according to the same general rule?*[11][12] We can see now that this approach fits the model quite well. The agent is considering the social well-being u_s in his personal utility function. If he or she thinks that everyone behaves as they do, when estimating the implication of other's actions with respect to social well-being, they are operating within a Kantian vision.

Naturally, this representation could be considered naive in face of reality. Why should each individual expect others to behave as he or she does? Most individuals do not, in fact, believe this. As noted in Brekke and Nyborg (2008), "...the categorical imperative defines one's moral responsibility vis-a-vis society without referring to others' *actual* behaviour, there is no presumption ... that he thinks others will *in fact* follow his example". Now, it is reasonable to say that everyone is different with respect to this (Kantian) moral responsibility. Some people are indeed more responsible than others. The weighting parameter α accounts for this fact. Larger values of α mean that the agent is being more responsible (Kantian) than homo-oeconomicus. Hence, those who care nothing for public well-being (or at least, do not behave as if they care) have $\alpha \approx 0$. Those who care (are responsible) and behave the best have $\alpha \approx 1$.[13]

[10] Recall that the pair (x, y) can be either $(0, 1)$ or $(\frac{1}{1+\rho}, 0)$.

[11] The original categorical imperative (or one of the original versions) was: "So act as if the maxim of your action were to become through your will a universal law of nature." Kant et al. (2002).

[12] For a well-written essay on the relationship between the Kantian imperative and climate change, see Rentmeester (2010).

[13] There is also another way of tackling this diversity: We might think that each person cares about social

Agent's Behaviour

Depending on his attitude toward the environment, concern and behaviour will emerge with different strengths. For example, someone with a stronger green attitude will have higher levels of concern for the environment and hence a greater response toward environmental conservation (or green behaviour). Following the model, the agent can choose to behave in a way that is green ($y = 0$) or grey ($y = 1$). In this case:

$$\begin{aligned} \text{Green (y=0):} \qquad & u(\tfrac{1}{1+\rho}) - \alpha \cdot d[(1-\delta)p_{t-1} + 0] \\ \text{Grey (y=1):} \qquad & u(1) - \alpha \cdot d[(1-\delta)p_{t-1} + \gamma] \end{aligned}$$

The agent will behave in a more green manner if the first term is greater than or equal to the second one. Applying this inequality to the previous equations and rearranging the terms, we get:

$$\alpha \underbrace{[d((1-\delta)p_{t-1} + \gamma) - d((1-\delta)p_{t-1})]}_{\Delta d \text{ : social cost of behaving grey}} \geq \underbrace{u(1) - u(\tfrac{1}{1+\rho})}_{\Delta u \text{ : cost of behaving green}} \tag{1.8}$$

$$\alpha \, \Delta d \geq \Delta u$$

At this point, two points are worth mentioning:

- $\Delta d(\gamma, \delta, p_{t-1}) = d[(1-\delta)p_{t-1} + \gamma] - d[(1-\delta)p_{t-1}]$ is increasing in p_{t-1} (since $d'' > 0$).
- $\Delta u(\rho) = u(1) - u(\frac{1}{1+\rho})$ is increasing in ρ.

The first observation implies that having higher perceived pollution levels, p^p_{t-1} (which is equal to p_{t-1}), will yield more people adopting green behaviour. This is simply because the condition in 1.8 is met for lower values of α when $\Delta d(\gamma, \delta, p_{t-1})$ increases. Therefore, a higher proportion of society will choose to behave in a green way. The second point corresponds to the obvious fact that the more expensive the green product is (higher values of ρ), the higher the cost (in terms of consumption) that will be borne by agents exhibiting green behaviour.[14]

well-being with the same intensity, but that the parameter α instead reflects how 'naive' (or optimistic) each person is. Although this is *not* the same idea stated here, a development and a proof of its equivalence is given in the Appendix 1.A.

[14] An interesting feature to note is that if we consider consumption levels as proportional to some income level w (i.e. comparing $u(w)$ and $u(\frac{w}{1+\rho})$), the cost of behaving in a green way, Δ_u, could be increasing, decreasing or independent of the income level w, depending on the functional form of $u(\cdot)$. Since in this formulation I explicitly leave aside the income effect, I use for simulations the case where $u(w) = \ln(w)$, which gives us a Δ_u that is independent of w. For details see Appendix 1.B.

Therefore we can see that for a given price of the green product and a perceived pollution level, there is a value of α^* that *divides* the society in two: those behaving in a green way ($\alpha \geq \alpha^*$) and those behaving in a grey one ($\alpha < \alpha^*$). Hence we can define a function $\theta(p_{t-1}, \rho)$ that tells us the proportion of people exhibiting grey behaviour for the values of p_{t-1} and ρ, as:[15]

$$\theta(p_{t-1}, \rho) = \min\left(\frac{u(1) - u(\frac{1}{1+\rho})}{d[(1-\delta)p_{t-1} + \gamma] - d[(1-\delta)p_{t-1}]}\ ,\ 1\right) = \min\left(\frac{\Delta u(\cdot)}{\Delta d(\cdot)}\ ,\ 1\right) \quad (1.9)$$

I note two important things about this new function $\theta(\cdot)$:

- There is a level of p_{t-1} ($p_{t-1_{min}}$) below which everyone's behaviour is grey. In other words, the environment is clean enough that no one 'cares' about it:

$$\text{Setting } \alpha = 1 \quad \rightarrow \quad d[(1-\delta)p_{t-1_{min}} + \gamma] - d[(1-\delta)p_{t-1_{min}}] = \Delta u$$

- There will be always some people exhibiting grey behaviour:

  We can always find $\alpha < \epsilon$, such that $\alpha\ \Delta d < \Delta u$, for any given $p_{t-1} > 0$ and $\rho > 0$. It is easily verified when using $\alpha = 0$: the agent has no incentive at all to behave in a green way.

We can graph this function with respect to p_{t-1}, for a given value of ρ, as in Figure 1.1.

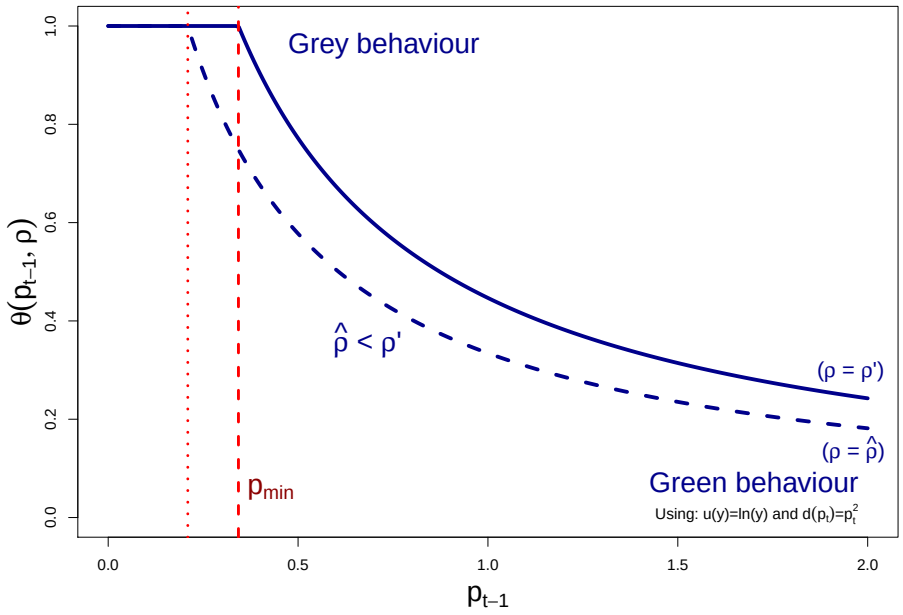


Figure 1.1: Share of grey behaving people w/r to perceived pollution level.

[15] For the present and following definitions, I use a uniform distribution of α. If this were not the case, we would have a different function $\theta(\cdot)$, but it would still retain the subsequent properties and results.

As we can observe via the dashed line in the illustration, a lower value of ρ (having a cheaper green product) will mean that more people adopt green behaviour. We can also notice that below a threshold level of pollution p_{min}, everyone behaves in a grey way: $\theta(p_{t-1}, \rho) = 1$. The intuition for this is quite straightforward: The environment is too clean to 'care' about, and therefore no one does. Alternatively, since the environment is so clean, it is simply too expensive to behave in a green way. We can also note that as the environment grows worse, the society becomes greener. This last observation comes from the fact that $\Delta d(\gamma, \delta, p_{t-1})$ is an increasing function of p_{t-1}. However, it is also intuitive. As the environment worsens, more people (those with less of a green attitude) become more aware and begin to prefer green behaviour.

It is important to note that the green behaviour is not coming from the agent thinking that his contribution to the environment will make things better (he assumes his contributions are negligible). The previous results comes from the idea that, as the (perceived) environment worsens, the agent's moral motivation increases with them (Δ_d) and therefore more people exhibit green behaviour. In the same fashion, if behaving green gets cheaper (lower values of ρ and hence lower Δ_u), and having the same moral motivation, more people will behave in a green way.

Now we can return to the pollution evolution (Equation 1.3). Rearranging the terms and defining the *change in pollution* as $\Delta p_t = p_t - p_{t-1}$, we have:

$$\Delta p_t = \underbrace{\gamma\, \theta(p_{t-1}, \rho)}_{\substack{\text{Actual} \\ \text{emissions}}} - \underbrace{\delta p_{t-1}}_{\substack{\text{Natural} \\ \text{absorption}}} \tag{1.10}$$

The first term corresponds to present emissions: It is the impact of consumption/production on the environment (γ) multiplied by the share of people behaving grey ($\theta(\cdot)$) multiplied by their grey consumption, which is 1. The second term is the natural absorption of the pollutant.

We graph these two terms in Figure 1.2.[16] As we can see from the figure, starting from a low level of pollution, $\Delta p_t > 0$, meaning that we will be polluting faster than what Mother Nature can absorb in the same period of time. At the beginning, for low levels of p_{t-1}, the society will behave in a way that is completely grey, $\theta(\cdot) = 1$, and from a point, p_{min}, we will see more and more people behaving in a way that is green (decreasing section of curve $\theta(\cdot)$). This curve will cross the straight line $\delta/\gamma\ p_{t-1}$ which represents the amount of pollution captured in a natural form. At this point, $\Delta p_t = 0$ means that the system stops evolving. It is easy to see that this equilibrium point is stable, since going further to the right will make $\Delta p_t < 0$.

[16] The figure has been rescaled by a factor of $1/\gamma$ in order to use the same previous Figure 1.1 and for simplicity in coming sections.

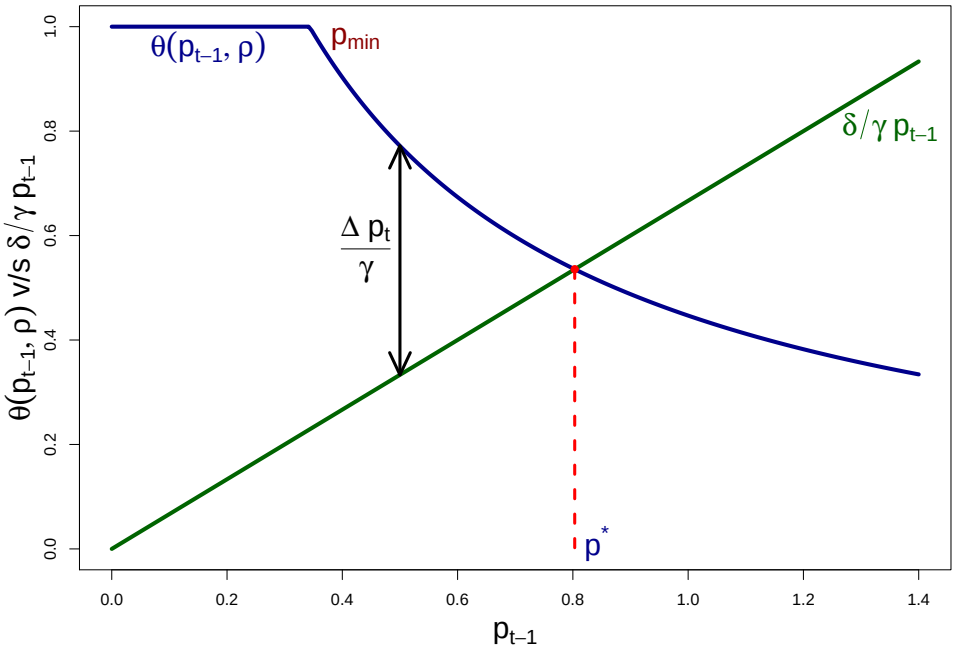


Figure 1.2: Evolution of pollution.

Let us now rotate the graph counter-clockwise, which will make it easier to analyse for further discussion. In doing so, we obtain Figure 1.3, which notes

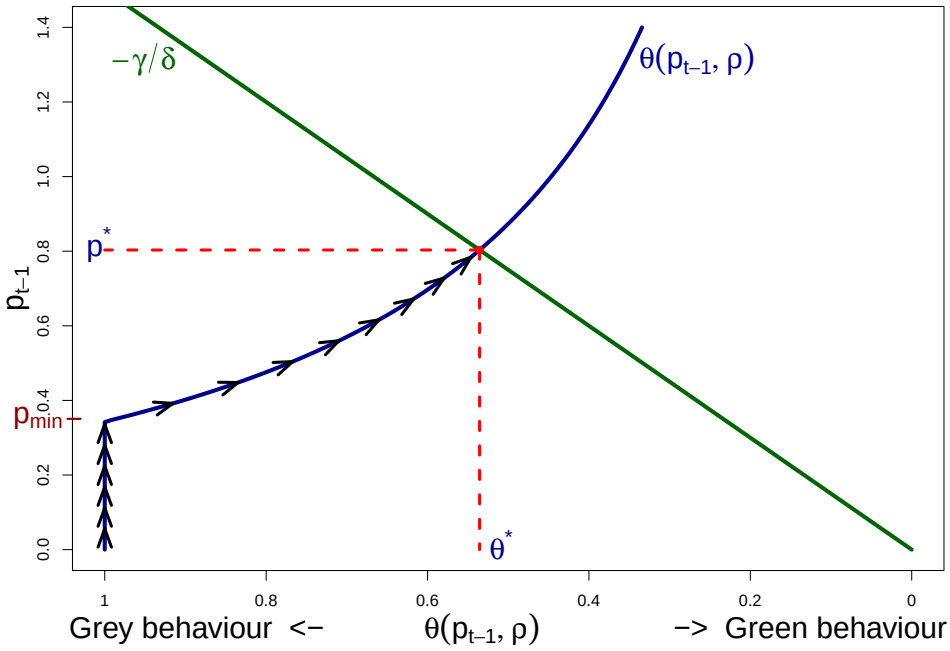


Figure 1.3: History of pollution.

the pollution level on the y-axis and how green the society is on the x-axis (the greenest being on the right side). The straight line (denoted by $-\gamma/\delta$) is again the natural absorption of pollution. The negative sign and change in terms comes from the rotational process. Since we know that when it crosses the curve $\theta(p_{t-1}, \rho)$ the system is at equilibrium, it is handy to leave it in the graph. In the same fashion, $\theta(p_{t-1}, \rho)$ in Figure 1.3 is the rotated version of $\theta(p_{t-1}, \rho)$ in Figure 1.2. Using an 'historical' approach, we begin from the lower-left corner and we then follow the arrows. At this starting point, there is no pollution and therefore everyone exhibits grey behaviour, $\theta(\cdot) = 1$ (recall that the x-axis is inverted). This leads to increasing levels of pollution up to the point where people begin to care about the issue, p_{min} (the kink to the right). As society continues to pollute, it becomes more aware of the issue and becomes greener. Society then reaches a point where emissions are equal to natural absorption, at (θ^*, p^*). At this point, a proportion $(1 - \theta^*)$ of the society is aware enough to exhibit green behaviour, which translates into an emission level equal to what Mother Nature can absorb at that pollution level p^*.

Endogenizing ρ: Introducing a political framework

I now introduce a simple political framework with two parties: the green party and the grey party. They only differ in that each party has a different environmental policy. For simplicity, I assume that the grey party does nothing about the environment, whereas the green party will implement a green-oriented policy. In this set-up, a green policy will simply be some tax/subsidy scheme to reduce the price gap between the green and grey products. This policy could be accomplished by taxing the grey products, subsidizing the green ones, or using both instruments at the same time. In other words, the green government will lower ρ, while the grey government will alter nothing. Other ways of making the green behaviour easier (cheaper) can be introduced by a green government. Recycle points could be closer to people's homes (which is not the case in most countries in the world), causing those who did not recycle previously (because it was too costly) to begin to to so. The idea is to reflect that having a green government will make green behaviour easier. In this simple model, it means lowering the cost of the green product by lowering parameter ρ.

It is implicitly assumed in this simple framework that as $\theta(p_{t-1}, \rho)$ decreases (and the society becomes greener), the government will implement a policy where ρ gets smaller. I base this assumption on an environmentally-ideological basis in the sense that green people will prefer green policies of this type. This idea also is supported by the results found by Comin and Rode in their paper "From green users to green voters" (2013). They show that as more German families started using photovoltaic panels, the Green Party received better results in the elections. But it is also a quite comprehensible assumption. Green people behave in a green way because they think it is the right thing to do (given

the current environment) and they will support a government that eases this behaviour.

Following the previous reasoning, we know that the elections will be won by the choice of the median voter. In other words, when the share of green people is bigger than some threshold $(1 - \overline{\theta})$ (and, hence, $\theta \leq \overline{\theta}$), a green government will be elected. In a simple case (for graphical illustration) where α is uniformly distributed, we have that $\overline{\theta} = 1/2$.

On the other hand, we have seen that if the value of ρ decreases, the function $\theta(p_{t-1}, \rho)$ will change, as in Figure 1.1. Now, since this policy is active only when $\theta(\cdot) \leq \overline{\theta}$ (on the right side of the graph), we have a $\theta(p_{t-1}, \rho)$ function with a discrete jump at $\overline{\theta}$, as in Figure 1.4.

We notice that we have two equilibria: θ_1^* a 'grey' equilibrium and θ_2^* a 'green' one. In the example depicted, $\theta_1^* < \overline{\theta}$. This gives us the possibility of two equilibria.

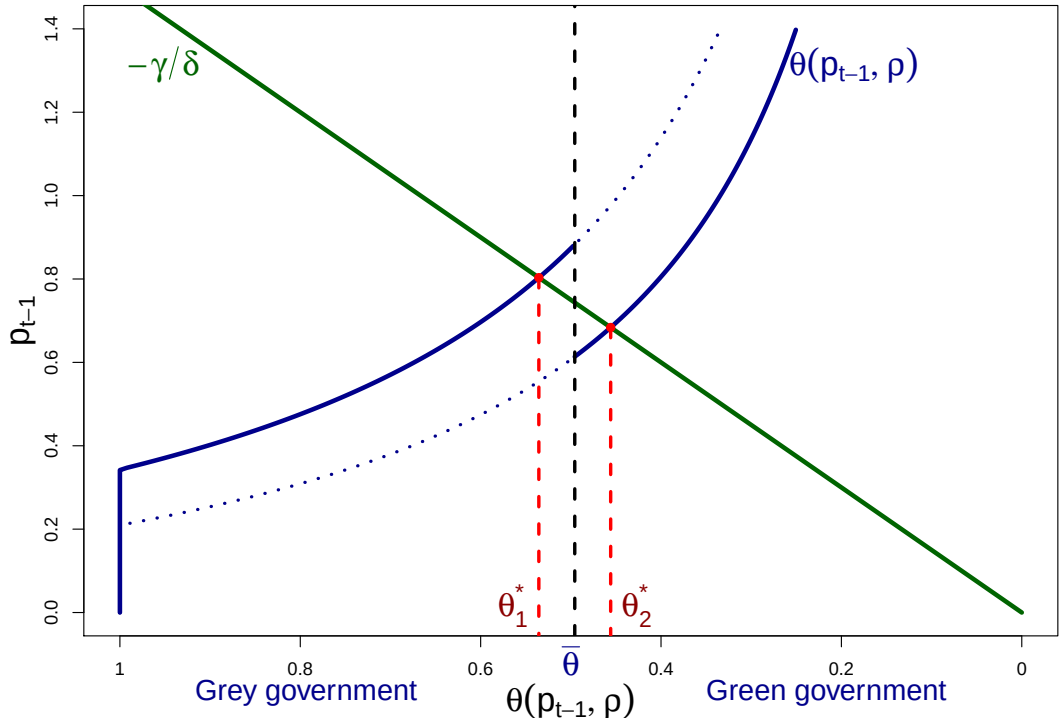


Figure 1.4: A green government.

It is interesting to note that the multiple equilibria occur *despite* the decrease in pollution levels. Having a better environment gives agents less moral motivation to exhibit green behaviour. But since there is a green government in place, behaving in a green way is also cheaper (the jump if $\theta(p_{t-1}, \rho)$ function), an effect that overrides the decrease in moral motivation. It could also be the case that some grey people (those close to $\overline{\theta}$) would

like to have a green government, even if their personal behaviour is grey. This could be due to the fact that these people feel the moral responsibility, but are unable or unwilling to bear the extra financial cost of behaving in a green way. Although this possibility might be true, I focus on the political model because it is simpler and, moreover, it is easy to realize that introducing this into the model would not change the dynamics of the society (we would simply see that everything shifts).

Now let us suppose that we are in a political framework with more parties involved (in the environmental spectrum). This assumption comes from the fact that, in the political arena, other topics are also involved in voters' (and politicians') decisions, such as income distribution, educational policies, etc. If this is the case, we can expect to have more 'jumps' in the dynamics, on the left and right side of $\overline{\theta}$. This could be due to:

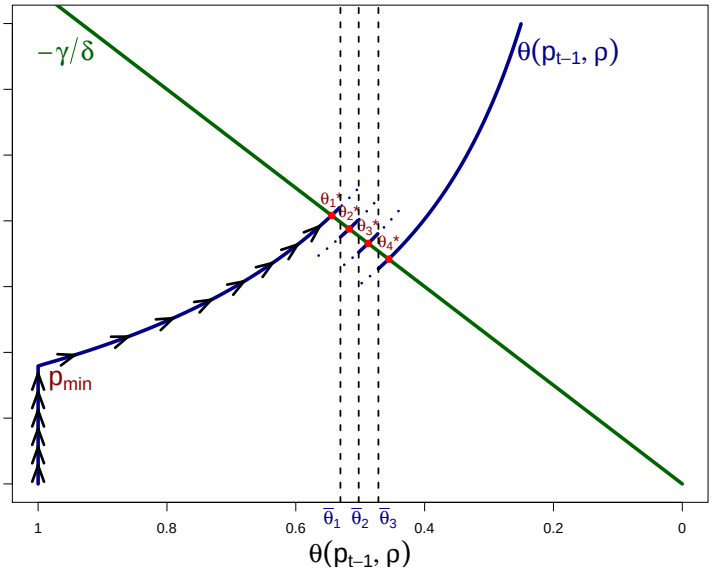
- The fact that coalitions might form in order to attract this 'multi-dimensional' median voter.
- A party or coalition might be willing to implement some green policy (a ρ level between the no policy and the full policy) in order to gain green voters. They are trying to attract voters who also share other political dimension(s) with this party or coalition.

If this is the case, we could have the following set-up:

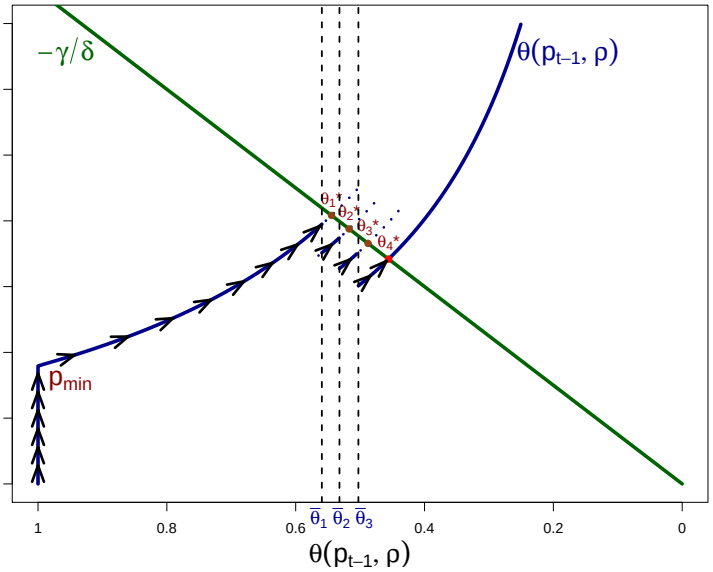
- $\theta > \overline{\theta}_1 \quad \rightarrow \rho_1$ (100% grey government)
- $\overline{\theta}_1 \geq \theta > \overline{\theta}_2 \quad \rightarrow \rho_2$ (partially grey government)
- $\overline{\theta}_2 \geq \theta > \overline{\theta}_3 \quad \rightarrow \rho_3$ (partially green government)
- $\overline{\theta}_3 \geq \theta \quad \rightarrow \rho_4$ (100% green government)

with $\overline{\theta}_1 > \overline{\theta}_2 > \overline{\theta}_3$ and $\rho_1 > \rho_2 > \rho_3 > \rho_4$. In this case, we can have different outcomes with $\overline{\theta}_1 \lesseqgtr \theta_1^*$, $\overline{\theta}_2 \lesseqgtr \theta_2^*$ and $\overline{\theta}_3 \lesseqgtr \theta_3^*$.

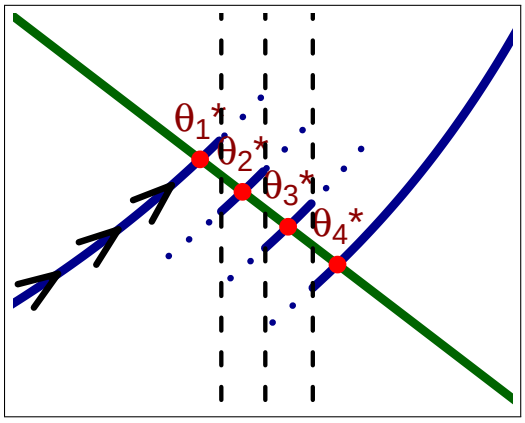
Two cases are depicted in Figures 1.5a and 1.5b:



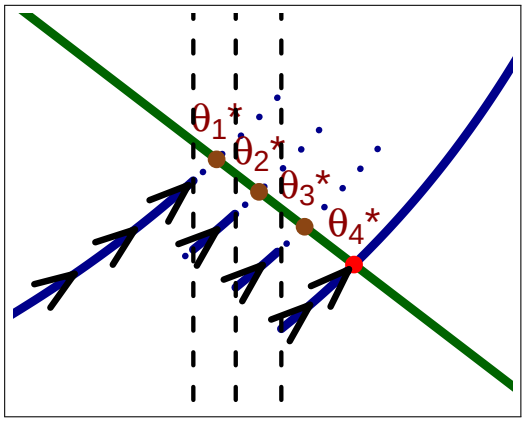
(a) $\theta_1 > \theta_1^*$, $\theta_2 > \theta_2^*$ and $\theta_3 > \theta_3^*$



(b) $\theta_1 < \theta_1^*$, $\theta_2 < \theta_2^*$ and $\theta_3 < \theta_3^*$



(c) Zoom of Figure 1.5a.



(d) Zoom of Figure 1.5b.

Figure 1.5: Different possibilities for multiple equilibria.

Observing the arrows drawn in Figure 1.5a, we can see how a natural evolution of the environment and society's behaviour could progress. As before, the system starts from the bottom left corner, where there is no pollution and everyone behaves in a grey way. Again, at some point (p_{min}), some people start to exhibit green behaviour and the system begins to shift to the right (always increasing). In this first case (Figures 1.5a and 1.5c), the system will arrive at equilibrium θ_1^*. However, if the conditions for Figure 1.5b (and 1.5d) are met ($\overline{\theta}_1 < \theta_1^*$, $\overline{\theta}_2 < \theta_2^*$ and $\overline{\theta}_3 < \theta_3^*$), the story becomes a different one. When moving toward θ_1^*, the system crosses the $\overline{\theta}_1$ threshold, jumping to a lower level of ρ (ρ_2) and not reaching the equilibrium θ_1^*. But it now moves toward θ_2^*. This process is repeated with $\overline{\theta}_2$ and $\overline{\theta}_3$ to finally arrive at equilibrium point θ_4^*. In other words, as the society becomes

greener in its behaviour, greener governments get elected. If the conditions are 'right', this might produce a cascading process that concludes with the (fully) green government in power.

However, the political framework might not be the only source of change in ρ. Another way of endogenizing ρ is from an evolution in the production process of the green products. As society becomes greener and more green products are bought, it would be logical to expect that: a) economies of scale begin to appear and; b) technological innovations occur in green product production methods (in contrast to a more mature grey production technology). Again, this last point could be 'artificially' induced by a political agreement in response to the society's demands. In the end, either method of endogenizing ρ leads to the same result: The system has a tipping point from which it moves toward a greener equilibrium.

1.3 A Green Nudge

As we saw in Figures 1.5a and 1.5b, all we need to nudge the system is to shift the threshold levels $\overline{\theta}_i$ or to move the curve $\theta(\cdot)$, even temporally. In this previous example, the two societies were not precisely the same, but I choose this framework in order to exemplify how a 'jump' from a grey equilibrium θ_1^* into a greener one θ_4^* might occur. To do so, we should recall that function $\theta(\cdot)$ depends on the price (differential) of the green product ρ as well as on the perceived pollution. This last term, as its name indicates, has to do with how people perceive the pollution levels, which I initially assumed to be correctly perceived ($p_{t-1}^p = p_{t-1}$ used in equation 1.4). Now we can assume that the agent has a perception bias Ω, giving us the following relationship: $p_{t-1}^p = \Omega\, p_{t-1}$. Changing the value of Ω will shift the curve $\theta(p_{t-1}, \rho)$ left or right. Hence, if we are able to change this bias, intentionally or by chance, then we could push the system from the grey equilibrium (or path) into the green one. As we expect, this variation does not need to be permanent. When the system has crossed the last threshold ($\overline{\theta}_3$), then there is no longer a need for such a nudge to sustain the new equilibrium.[17]

The change in Ω could be due to different causes. The perception of pollution levels depends on information. A higher level of exposure to information about climate change, which identifies the true environmental conditions, might alter this parameter. Extreme natural events, such as hurricanes hitting more frequently and severely (as with superstorm Sandy), could have the same impact (Lee and Markowitz (2013)). We have also seen how public attitudes dramatically changed after the Deepwater Horizon oil spill and

[17] We can see from this reasoning that it is not necessary to know exactly where the thresholds are, but only to be aware of the ability to nudge the system toward a greener equilibrium given these thresholds.

Fukushima Daiichi nuclear accident. These are not exactly the same phenomena modelled here, but the effect of external information on people's awareness and behaviour is the same. In the wake of the Fukushima Daiichi nuclear accident, the fear of nuclear disasters occurring elsewhere grew considerably, to the point where Germany permanently shut down eight of its reactors and pledged to close the rest by 2022. A similar sentiment arose in Italy, where more than 94% of voters opposed the government's plans to resume nuclear power generation in a June 2011 referendum. However, other major global players like United States did not change their policies in reaction to this event – not, at least, purely for that reason.

However, increasing information or changing perceptions are not the only ways of inducing change. Other avenues exist for nudging the system. One such channel could be political 'noise'. Specifically, a country could be close to its tipping point when a greener government is elected due to non-environmental reasons (left vs. right, social reforms, etc.). When this government implements green policy measures, this action could also trigger a cascading path into the green equilibrium. Independent studies by NGOs or the media might also yield the same type of nudge by increasing awareness about the environment and, hopefully, causing constituencies to push for changes to public policy.

In the same vein, it is worth noting that a multi-threshold situation meaning one in which there is more than one $\overline{\theta}_i$, eases the switching process. If we have a similar situation with more thresholds levels, centred in $1/2$, for example (meaning half on the $\theta(\cdot) < 1/2$ side and the other half on the $\theta(\cdot) > 1/2$ side), it follows that the cascading process begins at a lower level of greenness of the society (starting at higher levels of $\theta(\cdot)$). Also, it is easy to realise that the jump needed to switch regimes is bigger in the binary case compared to the more continuous one. In this sense, having a political arena where coalitions are more likely to form, could facilitate this switching process, increasing its likelihood of happening.

1.4 Social Approval and Social Pressure

So far I have used an *absolute* moral gain, meaning that the moral or green motivation of the agents comes only from an inner motivation. The agent acts in a given manner because they believe it is the right thing to do (the Kantian idea referred to above). I introduce now what I will call a *'relative'* moral gain. In this case, the agent also derives utility from being accepted by his peers and society in general. This idea has already been discussed in Hollander (1990), Nyborg et al. (2006) and Rege (2004). The concept is quite

simple: I get positive feedback from people behaving as I am behaving.[18] Therefore I call $u_a(\cdot)$ the satisfaction from *social approval* that agents get, which will in turn depend on his behaviour and on society's behaviour $\theta(\cdot)$. A corresponding weighting parameter β is introduced, producing the following version of the agent's utility function:[19]

$$U(\cdot) = u_p + \alpha\, u_s + \beta\, u_a \tag{1.11}$$

Concerning the form of u_a, I will assume that the agent gets a social reward from people behaving as he does.[20] In the interest of clearer notation, I will denote $\theta(\cdot)$ with θ_t, which is the share of grey people in the society at time t. The function $v(\cdot)$ will transform the share of people behaving as the agent behaves into social pressure, which is a strictly increasing function with $v(0) = 0$ and $v(1) = 1$ for normalization. In other words, this function translates the social behaviour θ_t into social pressure. Hence we have a piecewise function:

$$u_a(y, \theta_t) = \begin{cases} v(1-\theta_t) & \text{if } y = 0 \quad \text{(behaving green)} \\ v(\theta_t) & \text{if } y = 1 \quad \text{(behaving grey)} \end{cases} \tag{1.12}$$

We can immediately notice that if the society is mainly exhibiting grey behaviour ($\theta_t > 1/2$), the agent will *tend* to act in a grey way, and vice-versa. Recalling condition 1.8 for green behaviour and updating it to this new set-up, we have:

$$\alpha\, \Delta d(\gamma, \delta, p_{t-1}) \;+\; \underbrace{\beta\, [v(1-\theta_t) - v(\theta_t)]}_{\substack{\Delta a : \text{difference} \\ \text{in social approval}}} \;\geq\; \Delta_u(\rho) \tag{1.13}$$

In other words, when $\Delta_a > 0$, which occurs when $\theta_t < 1/2$, a bigger share of the society will behave in a green manner, and vice-versa.[21]

---

[18] It is interesting to note that this peer effect is also influenced by how 'public' our green actions are. In other words, we can declare that we behave in a quite green way, when maybe in reality we do not. A clear example of this can be found in Byrnes et al. (1999), which uses a real-life example from an electric utility green pricing program. Despite this fact, it is undoubtable that social pressure exists, especially considering the important and ever-present role of social networks in society and self-representation.

[19] It could be assumed that the agent interacts more with people behaving as he is doing, and therefore only has some probability of meeting with a random person, as in Bisin and Verdier (2000). It turns out that the resulting dynamics are the same, with a minor redefinition of the parameter β.

[20] I therefore disregard the case of negative social pressure from people *not* behaving as the agent does. It is easy to see, though, that including this second effect would not change the results.

[21] Since $v' > 0$, the point when $\Delta_a = 0$ (neutral social approval: $v(1-\theta_t) = v(\theta_t)$) has to be when $\theta_t = 1/2$. Hence, if $\theta_t < 1/2 \rightarrow \Delta_a > 0$ and vice-versa. In addition, this is a recursive way of defining the new function $\theta(\cdot)$. Another way of modelling this feature, would be to say that the agent reacts to θ_{t-1}, the share of grey people in the previous period. This would introduce a difference equation into the system, where lags would play a role too. In order to keep the model simple, I use the case where the agents

We now observe two things now:

- The minimum pollution where people start behaving green will be higher than the case without social pressure: $p'_{min} > p_{min}$

  If everyone is exhibiting grey behaviour and is affected by this peer effect (or *force*), the agent will have less incentive to behave in a green manner. In other words, the pollution will have to be higher in order for someone to care about it *and* endure this (new) social pressure. To find p'_{min}, we again set $\alpha = 1$ and solve: $\Delta_d(\gamma, \delta, p'_{min}) - \beta = \Delta_u$

- If the peer effect is too strong (β bigger than some threshold $\overline{\beta}$), the society will either behave in a completely green or grey way, switching at some pollution level $\overline{p}$ when $\beta = \overline{\beta}$.

  This follows the same intuition as the result obtained in Nyborg et al. (2006). They get a society with two extreme equilibria, completely green or completely grey, with a third unstable equilibria in between. The intuition is the following: Starting with the case in which everyone behaves in a grey way, when peer pressure is too strong, the pollution will increase and still no one will exhibit green behaviour. But at some (much) higher pollution level $\overline{p}$, some people, even bearing the high peer pressure, will start behaving in a green manner. In doing so, the peer pressure will be reduced and therefore more people will also behave in a green way. There will be a domino effect, with more people becoming green and reversing the peer effect towards being green, and therefore reaching a full green society. We notice that the new function $\theta(\cdot)$ is structurally different than its previous version, meaning that it cannot be worked out from the original set-up. For the proof of the last statement, see Appendix 1.C.

instantly responds to society's behaviour.

In order to make this idea clearer, I graph the changes obtained when peer pressure is present. Figure 1.6 is the new version of Figure 1.1 found on page 37.

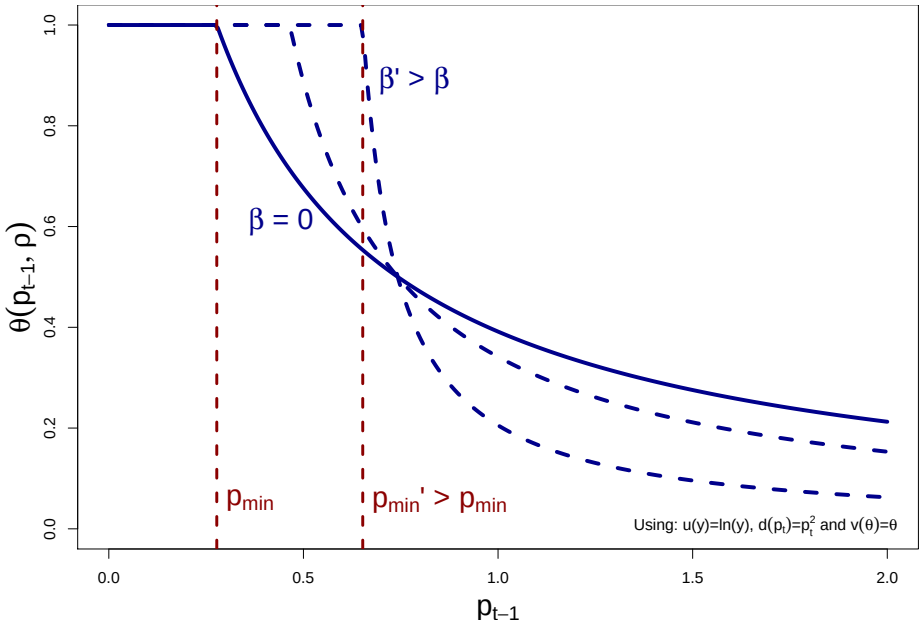


Figure 1.6: Share of grey behaving people with social approval.

First, we can notice that the three examples cross at a specific point on the graph. This feature is simply the fact that when $\theta_t = 1/2$, the peer effect disappears (it becomes neutral). Therefore, for any $\beta \leq \overline{\beta}$, the different curves should intersect at this point. We can use this fact to find the value of $\overline{p}$ mentioned before: It is just the pollution when $\theta(\overline{p}) = 1/2$, with the 'original' version of $\theta(\cdot)$. When $\beta > \overline{\beta}$, we will have a case with hysteresis. In this circumstance, the pollution level will have to be bigger than $\overline{p}$ in order to switch from grey to green. But once the society has switched to green, in order to switch back to grey the pollution level will have to be smaller than $\overline{p}$, hence the hysteresis effect. It is easy to note at this point, that the original version of $\theta(\cdot)$ and the actual one are structurally different.

Below I rotate the figure again in order to analyse how a society could evolve in the presence of peer pressure. As we can notice in Figure 1.7 (the updated version of Figure 1.4, found in page 41), the basic idea is the same. Now, however, the equilibrium points have moved apart from each other.

We can understand this effect as coming from some kind of *'attractors'* situated in each extreme ($\theta = 0$ and $\theta = 1$). Since in this particular case each equilibrium point is situated in either 'half' of the portrait ($\theta > 1/2$ and $\theta < 1/2$), they shift to the left and right side, respectively. Therefore, the resulting equilibria are more separated than before: The societies now behave dissimilarly. We can observe this in the two arrows drawn in the figure.

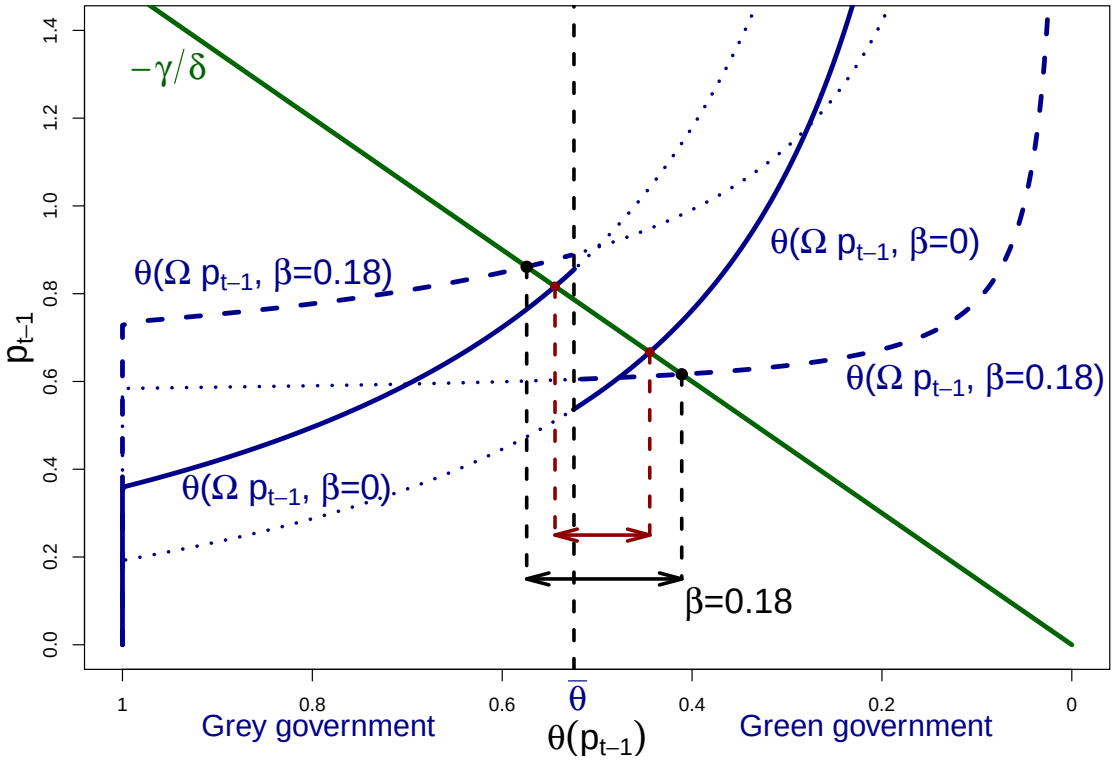


Figure 1.7: Two equilibria with social approval.

Recalling the cascading effect seen in the preceding pages, we notice now that this peer effect has detrimental consequences: The threshold needed to switch regimes is bigger ($\overline{\theta}$ has to be bigger), and at the grey equilibrium, the pollution is higher. For this reason it would be desirable (if our aim is to switch to the green side) to eliminate this grey peer effect. This idea may sound a bit too optimistic, but fortunately, history may be on our side. Thirty years ago, smoking was considered trendy. We were in the presence of the same effect: Being a smoker was fashionable, and others were encouraged to behave as smokers did. Today, this pro-smoking peer effect has almost completely disappeared or even become negative. Being a smoker is not considered good by anyone, even smokers! It is easy to see that a negative grey peer effect will have a positive outcome regarding the equilibrium and the chances of nudging the system.

Hence, if we can induce this negative grey social pressure (for example, with appropriate advertising), we could again induce this cascading process. Looking at the green side of the system, we might also note that the green peer effect is beneficial for society, at least in pollution terms, and it would drive society toward lower levels of contamination. A second effect also shows up: Being in presence of a green policy (lower levels of ρ) and having a peer effect ($\beta \gg 0$) pushes the system into a switching type. In other words, having a strong green social pressure and having cheap green products encourages a great

share of people to behave in a green fashion. We can observe this in the shape of the curve $\theta(p_{t-1}, \Omega, \beta = 0.18)$, which resembles a step function. We can again recall the smoking example: Smoking is now socially frowned upon.

1.5 Conclusions

This model attempts to explain why similar developed countries take different actions with respect to the environment. To do so, I model society as a composition of different types of people, from those who possess stronger attitudes toward the environment up to those who do not care at all about the environment. Green behaviour derives from being morally responsible. People with a stronger green attitude follow, up to some extent, a (green) Kantian imperative which makes them more prone to exhibit green behaviour. With this set-up, people can either behave in a green fashion (i.e., contribute to the environment) or in a grey way. The model reveals that the same society can arrive at two different equilibria, which might explain differences in environmental policies. This framework might also explain why countries with low income levels sometimes care about the environment more than their developed counterparts, as shown in Lee and Markowitz (2013).

Using the result of multiple equilibria, I have also shown that it is possible to switch from a grey trajectory to a green one. Providing information to people, for example, can raise their awareness about the environment and increase the chances of social change. This outcome reinforces the findings of Corbett and Durfee (2004) and Dunwoody (2007), who show that mass media has a large influence over social concerns, which can result in changes to a country's environmental behaviour and, eventually, its legal framework.

Is the nudge idea a concept that could be applied to explain different behaviours between Europe and the United States with respect to environmental issues such as carbon dioxide emissions? It might be the case that Europe's history of acid rain changed European perceptions and awareness about transnational pollution. When European countries faced acid rain caused by their neighbours' emissions, they became quite aware of the issue and established international environmental agreements, although only regarding these countries. This was not the case in North America, since acid rain was primarily resolved via nationwide actions of the US government. But this could be a bifurcation from an European grey path and could actually explain, up to some extent, the difference in behaviour between North America and Europe.

The present model does not try to provide a full explanation of this phenomenon, but rather to give some alternative insights of what could be occurring. The whole model

assumes that each person's green attitude is fixed. In other words, it is as if people were born with this trait and there is no chance of changing it over time. Obviously, this is an extreme assumption. However, the idea here is to create a model that can explain the faster changes in green behaviour observed over the last several years.

However, we could introduce a second slower evolutionary process of green attitude α (rather than using changes in behaviour directly). This process could be understood by considering how education, for example, changes attitudes over time. Such a change would be of the type described and modelled in Bisin and Verdier (2000). New generations approach environmental issues in a different way than their older cohorts, especially when they are taught about the environment (and the complications from environmental damage) from childhood. A possible empirical result of this effect is studied in Hersch and Viscusi (2005). The authors show that younger generations are greener than older generations. Adding this extension to the model would not change the previous results (especially the one concerning the nudge), and it would moreover reinforce the idea of how specific countries treat the environmental issue differently from a 'cultural' point of view. This last point could be a track for possible extensions of the present work.

It could also be interesting to explore the relationship between people's concern and their actions. In some cases it seems that the public is "concerned but unmoved," as stated by Oppenheimer and Todorov (2006). It seems as though this connection could be linked to people's values (Leiserowitz (2006)), which could be developed using a model with an evolution of attitudes.

Another extension might concern the relationship between the change in social awareness and/or behaviour and a more complex political framework model. A further line of work could be to verify this model with some empirical information. Unfortunately, some variables used in this set-up (for example, the concern of people and green attitude) are hard to properly measure. Even so, it would be a worthy venture to pursue.

Bibliography

James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving? *Economic Journal*, 100(401):464–77, June 1990.

S. Baumgärtner, T. Petersen, and J. Schiller. Bringing norms into action – the concept of responsibility. Working paper, University of Lünenburg, 2014.

Thomas Bernauer and Vally Koubi. On the political determinants of environmental quality. *annual meeting of the American Political Science Association, Hilton Chicago and the Palmer House Hilton, Chicago, Illinois*, 2004.

Alberto Bisin and Thierry Verdier. A model of cultural transmission, voting and political ideology. *European Journal of Political Economy*, 16(1):5–29, March 2000.

Kjell Arne Brekke and Karine Nyborg. Attracting responsible employees: Green production as labor market screening. *Resource and Energy Economics*, 30(4):509–526, December 2008.

Kjell Arne Brekke, Snorre Kverndokk, and Karine Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9-10):1967–1983, September 2003.

Annegrete Bruvoll, Bente Halvorsen, and Karine Nyborg. Household sorting of waste at source. *Economic Survey*, 4(2000):26–35, 2000.

Brian Byrnes, Clive Jones†, and Sandra Goodman‡. Contingent valuation and real economic commitments: Evidence from electric utility green pricing programmes. *Journal of Environmental Planning and Management*, 42(2):149–166, 1999.

Diego Comin and Johannes Rode. From green users to green voters. Working Paper 19219, National Bureau of Economic Research, July 2013.

Julia B Corbett and Jessica L Durfee. Testing public (un) certainty of science media representations of global warming. *Science Communication*, 26(2):129–151, 2004.

Sharon Dunwoody. The challenge of trying to make a difference using media messages. *Creating a climate for change*, pages 89–104, 2007.

Joni Hersch and W. Kip Viscusi. The generational divide in support for environmental policies: European evidence. NBER Working Papers 11859, National Bureau of Economic Research, Inc, December 2005.

Heinz Hollander. A social exchange approach to voluntary cooperation. *American Economic Review*, 80(5):1157–67, December 1990.

I. Kant, A.W. Wood, and J.J.B. Schneewind. *Groundwork for the Metaphysics of Morals*. Rethinking the Western Tradition. Yale University Press, 2002. ISBN 9780300094879.

Jon A Krosnick, Allyson L Holbrook, Laura Lowe, and Penny S Visser. The origins and consequences of democratic citizens' policy agendas: A study of popular concern about global warming. *Climatic change*, 77(1-2):7–43, 2006.

Timur Kuran. Now out of never: The element of surprise in the east european revolution of 1989. *World Politics*, 44:7–48, 10 1991. ISSN 1086-3338.

Jean-Jacques Laffont. Macroeconomic constraints, economic efficiency and ethics: An introduction to kantian economics. *Economica*, 42(168):430–37, November 1975.

Tien Ming Lee and Ezra Markowitz. Disparity in the predictors of public climate change awareness and risk perception worldwide. 2013.

Anthony Leiserowitz. Climate change risk perception and policy preferences: the role of affect, imagery, and values. *Climatic change*, 77(1-2):45–72, 2006.

Karine Nyborg. I don't want to hear about it: Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization*, 79(3):263–274, August 2011.

Karine Nyborg and Mari Rege. Does public policy crowd out private contributions to public goods. *Public Choice*, 115(3-4):397–418, June 2003.

Karine Nyborg, Richard B. Howarth, and Kjell Arne Brekke. Green consumers and public policy: On socially contingent moral motivation. *Resource and Energy Economics*, 28(4): 351–366, November 2006.

M. Oppenheimer and A. Todorov. Global warming: The psychology of long term risk. *Climatic Change*, 77(1-2):1–6, 2006. ISSN 0165-0009.

Torsten Persson, Gérard Roland, and Guido Tabellini. Comparative politics and public finance. *Journal of Political Economy*, 108(6):pp. 1121–1161, 2000. ISSN 00223808.

Gallup World Poll. `http://www.gallup.com/strategicconsulting/en-us/worldpoll.aspx`.

GlobeScan Radar. `http://www.globescan.com/expertise/trends/globescan-radar.html`.

Mari Rege. Social norms and private provision of public goods. *Journal of Public Economic Theory*, 6(1):65–77, 2004. ISSN 1467-9779.

Casey Rentmeester. A kantian look at climate change. *Essays in Philosophy*, 11(1):76–86, 2010.

John E Roemer. Kantian equilibrium. *The Scandinavian Journal of Economics*, 112(1):1–24, 2010.

Sarani Saha. Democratic institutions and provision of public good. University of California at Santa Barbara, Economics Working Paper Series qt55f3c17g, Department of Economics, UC Santa Barbara, April 2007.

Extreme Weather and Climate Change in the American Mind April 2013. http://environment.yale.edu/climate-communication/article/extreme-weather-public-opinion-April-2013.

Franz Wirl. Global warming with green and brown consumers. *Scandinavian Journal of Economics*, 113(4):866–884, December 2011.

Sammy Zahran, Samuel D Brody, Himanshu Grover, and Arnold Vedlitz. Climate change vulnerability and policy support. *Society and Natural Resources*, 19(9):771–789, 2006.

1.A 'Kantian' index and 'Naiveté' index equivalence.

As stated in the main section, there is also another way of tackling the diversity among agents. We might suggest that each person cares about social well-being with the same intensity, but that the parameter α instead reflects a person's naiveté or optimism. From this idea, we can use α as the share of individuals within society that each person thinks (or hopes) will behave as he does. I show here that this approach is equivalent to the approach outlined in the main section of the paper. In the main text, the original utility function is:

$$U(y) = u(y) - \alpha \cdot d[(1-\delta)p_{t-1} + \gamma \cdot y] \tag{A.1}$$

Now we might think of α as a 'naiveté' measure instead of a 'Kantian attitude' measure. This means that α will now reflect which proportion of the society the agent is expecting (or hoping) to behave as he does. In this new case, we get the following formulation:

$$U_2(y) = u(y) - d[(1-\delta)p_{t-1} + \alpha \cdot \gamma \cdot y] \tag{A.2}$$

Following again the same reasoning of the main section, I posit that the agent will behave in a green manner if $U_2(0) \geq U_2(1)$. We proceed again in the same fashion by rearranging terms and getting, for both cases (original and new), the following conditions for green behaviour:

$$\underbrace{\alpha\left[d((1-\delta)p_{t-1}+\gamma) - d((1-\delta)p_{t-1})\right]}_{\substack{\Delta d(\alpha,\gamma,\delta,p_{t-1}) \text{ : social cost of} \\ \text{behaving grey}}} \geq \underbrace{u(1) - u(\tfrac{1}{1+\rho})}_{\substack{\Delta u(\rho) \text{ : cost of} \\ \text{behaving green}}} \tag{A.3}$$

$$\underbrace{\left[d((1-\delta)p_{t-1}+\alpha\cdot\gamma) - d((1-\delta)p_{t-1})\right]}_{\substack{\Delta_2 d(\alpha,\gamma,\delta,p_{t-1}) \text{ : new version of} \\ \text{social cost}}} \geq \underbrace{u(1) - u(\tfrac{1}{1+\rho})}_{\substack{\Delta u(\rho) \text{ : same as} \\ \text{before}}} \tag{A.4}$$

We can again define an α^* that divides the society into those whose behaviour is green and grey. The only thing to do now is to check if this new function $\theta_2(\cdot)$ has the same properties as the original one $\theta(\cdot)$. Since the right hand sides of the inequalities are the same, I will only focus on the left hand sides. In the original version, we had the following properties:

$$\frac{\partial \Delta d(\alpha,\gamma,\delta,p_{t-1})}{\partial p_{t-1}} > 0 \qquad \frac{\partial \Delta d(\alpha,\gamma,\delta,p_{t-1})}{\partial \alpha} > 0 \qquad \frac{\partial^2 \Delta d(\alpha,\gamma,\delta,p_{t-1})}{\partial \alpha \;\; \partial p_{t-1}} > 0 \tag{A.5}$$

It is easy to verify that the same properties will hold for the case of $\Delta_2 d(\alpha,\gamma,\delta,p_{t-1})$. Hence we arrive at a new $\theta_2(\cdot)$ with the same properties of $\theta(\cdot)$ (although **not** the same function).

We can finally verify the two remarks made about $\theta(\cdot)$ for $\theta_2(\cdot)$:

- There is a level of $\Omega\, p_{t-1}$ $(\Omega(p_{t-1})_{min})$ below which everyone exhibits grey behaviour (the environment is clean enough such that no one 'cares' about it):

  Setting $\alpha = 1 \quad \rightarrow \quad d(\Omega(p_{t-1})_{min}) + \gamma) - d(\Omega(p_{t-1})_{min})) = \Delta u$

  This will actually gives us the **same** level of $\Omega(p_{t-1})_{min}$ as before.

- There will be always some people who exhibit grey behaviour:

  We can again find $\alpha < \epsilon$, such that $\Delta_2 d(\alpha, \Omega\, p_{t-1}) < \Delta u$, for any given $\Omega p_{t-1} > 0$ and $\rho > 0$. In the same manner, we can see that since $\Delta_2 d$ is continuous in α and that $\Delta_2 d(\alpha = 0, \Omega\, p_{t-1}) = 0$, then there exists an ϵ such that $\Delta_2 d(\epsilon, \Omega\, p_{t-1}) < \Delta u$ for given $\Omega p_{t-1} > 0$ and $\rho > 0$.

We can finally observe a graph showing both versions of the function $\theta(\cdot)$:

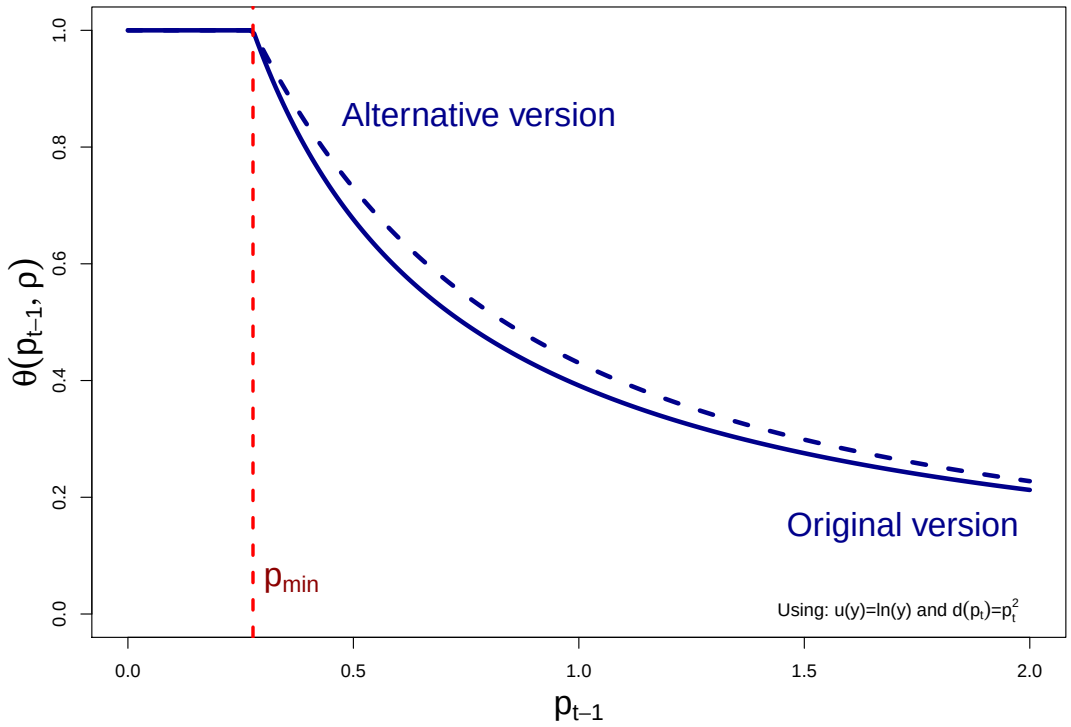


Figure 1.a: Alternative case where α is the agent's 'naiveness'.

A third and final option might be to suggest that both mechanisms (how Kantian each person is and how naive they are) are in place. For simplicity, I skip this alternative. However, if an individual's Kantian tendency is positively correlated to their optimism, the present results should hold.

1.B Cost of behaving green under other consumption utility functions.

Depending on the functional form of the consumption utility function, the cost of behaving in a green fashion Δ_u can be an increasing, decreasing, or constant function of the income level w. To observe this feature, I provide four examples using four different consumption utility functions.

$$u_1(c) = \ln(c) \tag{B.1}$$

$$u_2(c) = \sqrt{c} \tag{B.2}$$

$$u_3(c) = \frac{c-1}{c} \tag{B.3}$$

$$u_4(c) = \ln(c+K) \tag{B.4}$$

with K a positive constant.

Setting the grey and green consumption levels to w and $w/(1+\rho)$ respectively, we get:

$$\begin{aligned}\Delta_u^1 &= \ln(w) - \ln(w/(1+\rho)) \\ &= \ln\left(\frac{w}{w/(1+\rho)}\right) \\ \Delta_u^1 &= \ln(1+\rho)\end{aligned}$$

$$\begin{aligned}\Delta_u^3 &= \left(\frac{w-1}{w}\right) - \left(\frac{w/(1+\rho)-1}{w/(1+\rho)}\right) \\ \Delta_u^3 &= \frac{\rho}{w}\end{aligned}$$

$$\begin{aligned}\Delta_u^2 &= \sqrt{w} - \sqrt{w/(1+\rho)} \\ &= \sqrt{w}\left(1 - (\sqrt{1+\rho})^{-1}\right) \\ \Delta_u^2 &= \sqrt{w}\left(\frac{\sqrt{1+\rho}-1}{\sqrt{1+\rho}}\right)\end{aligned}$$

$$\begin{aligned}\Delta_u^4 &= \ln(w+K) - \ln\left(\frac{w}{1+\rho}+K\right) \\ \Delta_u^4 &= \ln\left(1+\frac{\rho}{1+\frac{K(1+\rho)}{w}}\right)\end{aligned}$$

We can observe that Δ_u^1 is constant, meaning it is independent of the value of w. In the second case, Δ_u^2 is increasing proportional to $\sqrt{w}$ (the term in parentheses is positive). In the third case, it is clear that Δ_u^3 is decreasing with w. The last case is a mixture between the first and second cases: Δ_u^4 is increasing with w but, as w grows, Δ_u^4 tends to a fixed value, $\ln(1+\rho)$, the same obtained in the first case.

1.C An (indeed) different dynamics with social approval.

As stated in section 1.4, adding a social approval gain to the agent's utility function does actually change (structurally) the function $\theta(\cdot)$, and therefore the dynamics. Adding social approval to the utility function will lead to a different functional form in the sense that the new $\theta(\cdot)$ function cannot be found by modifying the original parameters and/or by changing the (shape of the) consumption part $u(\cdot)$.

In order to prove this, we recall that the original function $\theta(\cdot)$ comes from Inequality 1.8 (page 36). There, I proved that for a given value of ρ and a pollution level p_{t-1}, there will always be some people exhibiting grey behaviour. I will prove now that this is not the case with social pressure, since it can be the case where for given values of ρ and p_{t-1} (and β), the society can become completely green. Recall the new condition for green behaviour, as in Inequality 1.13 (page 46):

$$\alpha\ \Delta_d(\gamma, \delta, p_{t-1})\ +\ \underbrace{\beta\ [v(1-\theta_t) - v(\theta_t)]}_{\substack{\Delta a : \text{difference} \\ \text{in social approval}}}\ \geq\ \Delta_u(\rho) \tag{C.1}$$

Let us verify whether this hypothesis of everyone behaving in a green manner ($\theta_t = 0$) is true. In this case, the social approval will be equal to: $\Delta_a(\theta_t = 0) = v(1) - v(0) = 1$. Therefore, the previous condition becomes:

$$\alpha\ \Delta_d(\gamma, \delta, p_{t-1})\ +\ \beta\ \geq\ \Delta_u(\rho)$$

Now, to verify that everyone is behaving in a green way, we can simply verify this condition for the least green person, the one purely motivated by economics, who has $\alpha = 0$. Therefore we get:

$$\beta\ \geq\ \Delta_u(\rho) \tag{C.2}$$

which is just the minimum weight of the social approval parameter in the agents' utility function. In other words, with this level of influence of peer pressure in agents' utility, even the pure homo-oeconomicus person bears the cost of green behaviour, only because of a social pressure source, and not because of a Kantian incentive, since he does not have one.

Chapter 2

Pushing the Tipping in International Environmental Agreements

2.1 Introduction

The present paper focuses on the idea that the creation of an International Environmental Agreement (IEA) that brings together many countries can be a colossal task to achieve. Barrett (1994), Barrett (2003), Rubio and Ulph (2006), and Eichner and Pethig (2013), among others, show that the number of signatories of self-enforcing IEAs generally does not exceed three or four; or if it does, their emissions are almost the same as business as usual (BAU). Recent failures to reach an effective agreement concerning climate change are a clear example of this fact. Some solutions that have been analysed to arrive at a successful IEA involve the idea of starting with a coalition composed of a small sub-set of countries and then incorporating more countries. A first possibility explored by Heal and Kunreuther (2011) suggests enlarging a small coalition in a cascading process. They assume a positive reinforcing effect among signatories, which induces the accession of more countries. In a similar manner, Barrett (2006) shows that if there is a green technology that exhibits increasing returns to adoption, a tipping coalition size exists after which countries join to finally reach the grand coalition. A second line of thought is analysed by Carraro and Siniscalco (1993), Hoel and Schneider (1997), and Barrett (2001), where transfers are used in order to induce more countries to join the IEA. The first paper shows that a commitment problem prevents the formation of the grand coalition – or if it forms, the countries still function as though it were business as usual (BAU). The commitment problem comes from the fact that the IEA set-up resembles a chicken game: Signatories are better off compared to not having an IEA at all, but they would prefer to have others to sign and abstain themselves. Hence, when transfers are made in order to enlarge the coalition, the original signatories prefer to leave. Hoel & Schneider (1997) show a similar effect, where commitment is gained by conformity. Barrett (2001) finds a

similar result as Carraro & Siniscalco (1993) but with asymmetric countries. If asymmetry is weak, the commitment issue arises. If there is strong asymmetry, transfers can sustain a superior outcome compared to an IEA without side payments.

Given all these barriers, I explore a new tactic. I show that using a border tax, a technological transfer, or both can effectively induce and sustain the grand and meaningful coalition.[(1)] I find the conditions under which each case is optimal. It turns out that there is no general recipe, although at least one instrument is always better than nothing. I also show that if transfers are part of the optimal solution, they have to be made to the least-green countries. This choice improves the chances of success while minimizing the amount of transfers. These results could be quite important in face of the recent failures in reaching an IEA that effectively tackles climate change. It is possible that the reasonings presented in this paper could light the ongoing negotiations and offer potential solutions.

In order to develop this idea, I build a model that is an expanded version of Barrett (1997). In his paper, the world is composed of identical countries, each containing one firm that produces an homogeneous and traded good and pollution as a side product. In Barrett's three-stage game, each country decides to join an IEA or not (one shot). In the second stage, the formed coalition is group rational and plays accordingly, with non-signatories playing as singletons. Finally, firms maximize their profits by choosing their segmented outputs à la Cournot, and markets clear.

Starting from this set-up, I first add asymmetries among countries in terms of damages from emissions and abatement costs. These asymmetries are empirically motivated. Societies exhibit green behaviour, even though direct 'rational' thinking could command otherwise. This could be related to moral reasons rather than economic ones. As shown in Chapter 1, structurally similar countries (in terms of development level and political system) can end up behaving differently with respect to global environment. This could be for historical reasons, but the point is that some countries have become (quite) aware of the environmental problem at hand and have started acting accordingly. Hence, differences among countries are how green they are and what abatement technology they have in place. With respect to the first point, an equivalent interpretation could be how countries are affected by pollution. Therefore, having higher marginal damages coming from emission will be treated as a synonym of being greener. On the abatement side, countries can either have bad technology (high marginal cost of abatement) or good technology (low marginal cost of abatement). Conversely, it can be thought that historically green countries have invested in cleaner technologies (to produce electricity, for instance),

[(1)] Meaningful in the sense that countries are actually making an effort in abating and not simply acting in a way similar to BAU.

which is equivalent to having technology with a cheap abatement cost. It looks like Europe has followed this path, by investing in green technology and auto-imposing self-constraints on emissions. Following this, I assume that there are two groups of countries: rich countries, with high environmental concern and low cost abatement technology; and poor countries (also called outsiders), with lower environmental concern and expensive abatement technology. Also, I assume that for rich countries it is profitable to form a coalition, where it is not for poor ones.

Second, I add the possibility of transfers between countries, which creates the potential to enlarge the existing coalition in a fashion similar to Carraro and Siniscalco (1993). However, I consider technology transfers (instead of monetary transfers, as Carraro and Siniscalco do), which can change the game itself. The final addition to the baseline model is that I allow the coalition to impose a border tax on non-signatories. As expected, the idea behind this is to deter free-riding (and possibly induce accession), similarly to the intention of the trade ban in Barrett's (1997) paper.

Through this set-up, I want to study if a small initial coalition of green countries can induce the formation of the grand coalition. The idea behind this is the same as the successful mechanism implemented in the Montreal protocol and its subsequent amendments (for a detailed explanation, see Barrett's book (2003)). In this case one country, the United States, was unilaterally willing to cut emissions and consumption of ozone-depleting substances. They were well aware that at first, leakage coming from trade could reduce their effort results, but knew that if other big economies would follow suit, the gains (both nationally and globally) would be much larger. Therefore, they were willing not only to unilaterally ban the use and production of CFCs, but also to ban trade of these substances. This second component, plus the fact that they were also willing to help developing countries to switch to new and clean substances, pushed other countries to join what would become one of the most successful IEAs in history.

Obviously, it would be appealing to do a similar thing with greenhouse gases (GHG). However, while tackling the production and trade of CFCs is a simple task, it is quite impossible for CO_2 emissions, since they are embedded in almost every product we trade. Of course, completely banning trade seems out of the question. The 'sticks' used in Montreal are not credible in an IEA concerning CO_2. To avoid this obstacle, we can use a border tax, imposed on goods coming from non-signatories countries, in order to deter free-riding and induce accession. This will be the case in the present paper, acknowledging that this could bring on a trade war. With respect to this, I assume that coalition members can impose a border tax and non-signatories do not retaliate. I then drop this

assumption and analyse how credible this is and which cases it actually holds true for.[2] Furthermore, in a recent paper by Nordhaus (2015), which studies a DICE model with 15 specific regions, the author also assumes that it is possible to impose a border tax without retaliation and shows that this tax induces countries to join the coalition. I find a similar result, plus an upper limit for this tax imposed by the condition of inducing the grand coalition.

With all this in mind, I analyse if a border tax imposed by signatories, a technological transfer, or both can induce the grand coalition. To do so, I study the stability of the grand coalition for different parameter scenarios.[3] I analyse the different cases in which either the tax, the transfer, or both together sustain the grand coalition. These solutions depend on, among others things, the countries' environmental damage (or concern). This result fits with the aforementioned paper, since Nordhaus also finds mixed effects. Then I assume that both instruments are needed and I find the optimal recipient group. I give an example of this case, which portrays how both instruments work together.

Finally, I investigate the case in which we can discard the Minimum Participation Clause (MPC).[4] These clauses are legal tools that help IEAs reach the desired equilibrium as in, for example, the Montreal case. However, in this particular case, the MPC was not too high.[5] Concerning the CO_2 problem, although there is a consensus about the damages arising from global warming, there are differences in opinion among countries regarding how these damages will hurt them specifically. Moreover, the damage level is uncertain and eventually comparable to the cost of switching to the required clean technology. Uncertainty about the gains and costs of such an agreement makes the political decision a much harder one to pursue and puts the MPC on a level that might not be politically feasible to reach. Therefore, I perform a theoretical exercise by verifying what would be needed in order to reach the grand coalition without an MPC. The goal is to illustrate the mechanics of what would have to be done to lower the necessary MPC.

The rest of the chapter is structured in the following way: Section 2.2 presents the

[2] The results found here are also in line with those of Anouliés (forthcoming), where the author shows that for some parameter configurations there is no retaliation.

[3] In order to study a meaningful set of parameters, I assume that parameters are such that in all possible cases countries are trading with each other. Also, parameters' space is delimited in order to have a stable coalition of rich countries.

[4] The Minimum Participation Clause states that the treaty becomes binding when an amount of countries greater than or equal to this minimum have signed the agreement. Since IEAs are faced with multiple equilibria (typically two), in order to induce the optimal one, treaties include an MPC. The objective is to make sure that signatories do not suffer when few countries have signed the agreement.

[5] In order for the Montreal protocol to become binding for parties, it would have to be ratified by at least 11 countries accounting for at least two-thirds of the 1986 level of global consumption of the controlled substances.

model and solves the firm maximization program. Section 2.3 analyses if a border tax, a technological transfer, or both induce the formation of the grand coalition. Section 2.4 finds the optimal recipient group and analyses the special case with no MPC. Section 2.5 concludes.

2.2 The model

I develop a two-stage game built on Barrett (1997). In the first stage, each country chooses whether or not to be part of the coalition of size k, S_k. There are N countries that are asymmetric in two dimensions. First, the technology they have access to can offer a low marginal cost of abatement (σ_L, good technology) or a high one (σ_H, bad technology). Second, countries differ in terms of the marginal damage from emissions they suffer from, noted ω_j, for country j. I will be using marginal damage from emissions and environmental concern as synonyms; in this line, I will also be referring to countries with higher environmental concern as 'greener'. I then define two types of countries. The 'rich' are those with the good abatement technology, σ_L, and a high marginal damage from emissions, ω_H. The 'outsiders' or 'poor' have the bad abatement technology, σ_H,and a lower environmental concern than the rich: $\omega_j < \omega_H$, where ω_j, the marginal damage from emissions of country j, is a continuous variable.

Departing from Barrett's (1997) model, I assume that if a country enters the coalition, it will fully abate ($q_j = 1$). If not, it will abate nothing ($q_j = 0$). Hence, for country j, joining the coalition and having $q_j = 1$ become synonyms.[6] In the non-signatory case, this assumption poses no real restriction, since $q_j = 0$ is optimal for these countries (using some standard assumptions), which is not the case for signatories. The assumption can be understood as follows: countries choose between two types of technology – clean and dirty. The clean one does not pollute at all, which translates in this model into full abatement; the dirty technology pollutes, which is equivalent to no abatement.[7] In the present framework, this is equivalent to a σ_L below some threshold and σ_H above some other threshold.[8] Also, this feature has two good implications: First, I discard by construction the case of having coalitions (especially the grand coalition) that do not abate, or who operate quite close to BAU when they do abate, as the literature has shown (as in Barrett (1994) and Eichner and Pethig (2013)). Second, it makes the model more tractable with clear-cut results.[9] On another side, if countries' decisions are binary, it makes the

[6] Conversely, for country j, not joining the coalition and having $q_j = 0$ are also synonyms.

[7] Alternatively, it can assumed that there are two emissions levels – low and high. It is easy to see that this framework is equivalent to the present one, in which I have 'normalized' the low emission level to zero.

[8] These thresholds are calculated explicitly in Section 2.3.

[9] Binary choices can also be found in the literature, as for example, in Barrett (2003) and Heal (1994).

coalition formation a harder process, since becoming a member of the IEA implies full abatement for the joining country. Finally, signatories can not punish a country for leaving the coalition by increasing emissions.

Focusing on abatement technology rather than on emissions is important: From a political point of view, it can more easily lead to an enforceable IEA since the technology being used is more easily verifiable than total emissions. Furthermore, it allows us to consider technology transfers, which change the recipient country's incentives to join the coalition, and therefore the game itself, which is part of the overall plan.

Countries inside the coalition tax imports, at a rate t, of goods produced in non-signatory countries. The main obstacle to reaching a meaningful IEA comes from having carbon leakage, because it provides countries strong incentives to free-ride. The border tax is, then, a credible tool for hindering leakage. In this set-up I assume that only signatories tax goods coming from non-signatories and that the latter do not retaliate with another tax on signatories' goods. First, general results will be derived using this framework and, in the following subsection, I drop this assumption and analyse if non-signatories prefer to retaliate.

In the second stage, firms move by choosing simultaneously their segmented outputs, all within a Cournot-Nash set-up. It is a perfect information model in the sense that countries perfectly know their costs and gains, as well as those of other countries. Focusing now on the solution of the game at hand, I proceed using backward induction.

2.2.1 Firms' choices

There are N countries and N firms (one per country) that produce an homogeneous traded good and a transboundary pollution. The inverse demand in each country j is given by $p\left(x^j\right) = 1 - x^j$, where x^j is consumption in country j. Costs of the firm in country j are $C\left(\sigma_j, x_j, q_j\right) = \sigma_j q_j x_j$, where x_j is the total output of the firm, $\sigma_j \in \{\sigma_L, \sigma_H\}$ is its marginal abatement cost, and $q_j \in \{0, 1\}$ is the abatement standard chosen by the government of country j, taken as given for the firm in this country. Emissions by firm j are $x_j\left(1 - q_j\right)$: if abatement is maximal, emissions are zero, and if no abatement is undertaken, emissions are equal to output. The marginal abatement cost of firm j could reflect the technology used – to produce electricity, for example – in country j. Therefore it could be thought as the (accumulative) efforts undertaken by a country to be greener.

Firms choose their output for each market simultaneously. Transport costs are zero, and each firm takes its own abatement standard and the segmented outputs of other firms as given. Firm j chooses a quantity to produce and ship to market i, x_j^i, so as to maximize

its profit π_j:

$$\max_{x_j^i \geq 0} \pi_j = \sum_{i=1}^{N} (1 - x^i - t_j^i - \sigma_j q_j) x_j^i \tag{2.1}$$

with $t_j^i = t$ if $i \in \mathcal{S}_k \wedge j \notin \mathcal{S}_k$, and $t_j^i = 0$ otherwise. There are N first order conditions for firm j:[10]

$$1 - x^i - t_j^i - \sigma_j q_j - x_j^i = 0 \qquad \forall i, j \tag{2.2}$$

Taking into account the fact that $q_j \in \{0, 1\}$ and solving the system of equations formed by the N first order conditions for the N firms, we get:

$$x^i = \begin{cases} \frac{N - \sigma_{\mathcal{S}} - (N-k)\cdot t}{N+1} & \text{if } i \in \mathcal{S}_k \\ \frac{N - \sigma_{\mathcal{S}}}{N+1} & \text{if } i \notin \mathcal{S}_k \end{cases} \tag{2.3}$$

where k is the size of the coalition, and $\sigma_{\mathcal{S}} = \sum_{j \in \mathcal{S}} \sigma_j$ (the sum of the coalition's marginal abatement costs). Replacing this result into the FOCs, we have:

$$x_j^i = \begin{cases} \frac{1 + \sigma_{\mathcal{S}} - (N+1)\sigma_j + (N-k)\cdot t}{N+1} & \frac{1 + \sigma_{\mathcal{S}} - (k+1)\cdot t}{N+1} & \text{if } i \in \mathcal{S}_k \\ \frac{1 + \sigma_{\mathcal{S}} - (N+1)\sigma_j}{N+1} & \frac{1 + \sigma_{\mathcal{S}}}{N+1} & \text{if } i \notin \mathcal{S}_k \end{cases} \tag{2.4}$$

$$\begin{matrix} \uparrow & \uparrow \\ \text{if } j \in \mathcal{S}_k & \text{if } j \notin \mathcal{S}_k \end{matrix}$$

From where we can calculate the firm profit π_j and production x_j:[11]

$$\pi_j = \begin{cases} \frac{N \cdot \aleph_j^2 + k(N-k)t\left(2\aleph_j + (N-k)t\right)}{(N+1)^2} & \text{if } j \in \mathcal{S}_k \\ \frac{N \cdot \aleph_2^2 + k(k+1)t\left((k+1)t - 2\aleph_2\right)}{(N+1)^2} & \text{if } j \notin \mathcal{S}_k \end{cases} \quad \text{with} \quad \begin{aligned} \aleph_j &= 1 + \sigma_{\mathcal{S}} - (N+1)\sigma_j \\ \aleph_2 &= 1 + \sigma_{\mathcal{S}} \end{aligned} \tag{2.5}$$

$$x_j = \begin{cases} \frac{N \cdot \aleph_j + k(N-k)t}{N+1} & \text{if } j \in \mathcal{S}_k \\ \frac{N \cdot \aleph_2 - k(k+1)t}{N+1} & \text{if } j \notin \mathcal{S}_k \end{cases} \tag{2.6}$$

2.2.2 Countries' choices

Country j's net benefits are the sum of firm j's profits, plus the surplus of country j's consumers, less the environmental damage suffered, plus border taxes collected, if

[10] I only consider those situations where firms produce positive quantities in equilibrium.

[11] Mathematical development of equations (2.2), (2.3) and (2.4) are in Appendix 2.A.

that is the case. Given demand specifications, the consumer surplus is equal to $(x^j)^2/2$. Pollution is assumed to be a pure public bad, and aggregate emissions are given by $\sum_{i=1}^{N} x_i(1-q_i)$, where x_i is the total output of firm i. The constant marginal environmental damage is equal to ω_H for the rich countries, and to $\omega_j < \omega_H$ (abusing the notation) for a country j belonging to the group of *outsiders*. If country j is a signatory country (meaning that $q_j = 1$), the border taxes collected are equal to the tax rate t, times the imports from non-signatories (all countries i such that $q_i = 0$). With these, country j's profit is:

$$\Pi_j = \pi_j + (x^j)^2/2 - \omega_j \left[\sum_{i=1}^{N} x_i(1-q_i) \right] + t \cdot q_j \cdot \sum_{i=1}^{N} x_i^j(1-q_i) \tag{2.7}$$

where π_j, x^j, x_i and x_j^i are expressed in terms of k and model parameters (given by equations (2.3) and (2.4)). Countries choose whether or not to join the coalition, and therefore whether to fully abate or not at all, by maximizing this profit.

2.2.3 Coalition formation

Countries decide either to join the coalition or stay outside of it, in which case they act as singletons. Countries join the coalition if it is profitable for them to do so; they decide this by comparing their profit when belonging to a coalition of size k to their profit when staying outside it (resulting in a coalition of size $(k-1)$). Being profitable can have two meanings. We can either assume that it is profitable for each country, individually and *without* profit sharing among the coalition, or we can assume that the countries belonging a coalition share their profits according to some sharing rule. Since the second case is better for the coalition compared to the individual option, I assume that the coalition forms according to this criteria. Nevertheless, noting that country profit (Eqn. 2.7) depends linearly on ω_j, some analyses are made using an individualistic approach and then extended to the one where signatories share gains.

In order to analyse the coalition stability, I rely on the well-known concepts of Internal Stability (IS) and External Stability (ES)[(12)]. In the case of the grand coalition, we only need to verify for IS, i.e. checking that the sum of profits of coalition members is greater than the sum of their *outside options* (meaning the profit each country would make if it were to leave the coalition of k countries and a coalition of the remaining $(k-1)$ countries

[(12)]Internal and External Stability as defined by D'Aspremont et al. (1983) and subsequently used by a substantial literature. IS and ES mean that for a given coalition, no member prefers to leave and no non-signatory wants to join, respectively.

formed). The IS condition can be represented as

$$\sum_{j \in \mathcal{S}_k} \Pi_j^s \; > \; \sum_{j \in \mathcal{S}_k} \Pi_j^{oo} \tag{2.8}$$

where Π_j^s is the country profit of a signatory (belonging to the coalition of size k: $\mathcal{S}_k$) and Π_j^{oo} is the country outside option (for the same group of countries). If inequality (2.8) is divided by k (the coalition size), we can talk of the mean coalition profit and the mean outside option, which will turn out to be more convenient.

I also assume that the parameters of the model are such that there exists a small coalition of rich countries, which is stable (i.e. IS and ES). The group of rich countries is denoted by $\mathcal{D}$. This coalition can be expanded, especially to try to induce the grand coalition, by the use of a border tax and/or technological transfers. The latter means that some r countries can receive a technological transfer that reduces their marginal abatement cost from σ_H to σ_L. This recipient group is denoted by $\mathcal{R}$. These technological transfers can be thought of as international aid from rich countries to some other countries in order to shift from carbon-intensive power sources toward more eco-friendly ones – for example, to change how electricity is produced. This one-time transfer costs K per recipient.

2.3 Inducing the grand coalition

As stated in the Introduction, I want to verify if using two instruments is better than using one or none in order to induce the grand coalition. The intuition for this goes as follows: technological transfers (TT) lower costs. However, if benefits are low, poor countries will not come in. A border tax (BT) raises the cost of being out, but if costs of abatement are high and environmental concern for these countries is low, poor countries might prefer to be out even when facing the BT. So what could matter is the combination of a BT and TT.

Following this idea, I analyse four situations: Case (1) An IEA with no BT and no TT; Case (2) IEA with BT but no TT; Case (3) IEA with TT but no BT; and Case (4) IEA with BT and TT. With all this, I focus on conditions needed for the grand coalition (GC) to be Internally Stable (IS) in these four cases.

Therefore, the idea is to analyse if the use of one or two instruments enlarges the parameter space within which the grand coalition is stable. I am interested in seeing in which condition using both instruments (Case (4)) works better, and the same idea for Cases (2) and (3). In this context, to 'work better' means that the parameter space in which the GC holds is enlarged. In other words, it means that using one or two instru-

ment makes the GC stable, where this was not the case without the help of these instruments. At the same time, it means that those parameters that already sustained the GC still do, but with greater profits for joining the GC (the coalition gets 'stronger').

As mentioned in section 2.2.1, parameters are such that the firm's program has interior solutions, which translates into $x_j^i > 0$. On the other hand, and as mentioned in the previous section, I explore those cases (parameter values) where we are in the presence of a (small) initial coalition composed of only rich countries. This means that rich countries are sufficiently green and enjoy cheap technology (or some combination of both). At the same time, this means that outsiders or poor countries care so little about the environment and/or their abatement technology is so expensive that they prefer to stay out of the coalition. Consequently, we have the following conditions:

1. In any scenario, there are interior solutions for the firm's maximization program. Specifically, x_j^i (production in firm j being shipped to country i) has to be greater than zero.

2. Rich countries want to abate when all other rich countries are abating.

3. Poor countries, knowing that d rich are abating, prefer not to abate.

These last two conditions mean that a coalition of d rich countries is Internally Stable (IS) and Externally Stable (ES). Condition 1 translates into various inequalities, depending on which country is exporting to which other country. These inequalities limit parameters σ_L and σ_H as well as pose limits into the border tax level t. Applying these constraints, we get:[13]

$$\sigma_L < \frac{1}{N-d+1} = \sigma_L^{max} \tag{2.9}$$

$$\sigma_H < \frac{1+d\sigma_L}{N} \tag{2.10}$$

$$\sigma_H < \frac{N+1}{N(N-d+1)} \tag{2.11}$$

$$t < \frac{1+d\sigma_L+(N-d-2)\sigma_H}{N-1} \tag{2.12}$$

[13] Limits for σ_L (Eqn. (2.9)) and t (Eqn. (2.12)) come from $x_j^i > 0$ in combination to the FOC in Eqn. (2.2). This is verified for different cases: rich country exporting to a poor country, rich to rich, etc. Condition (2.10) comes from $\sigma_L < \sigma_H$. Condition (2.11) is just the combination of Condition (2.9) and Condition (2.10). These conditions have to hold for any coalition size. I have assumed d is much smaller than N, in this case $d < N - 3$.

Conditions 2 and 3 delimit ω_j and ω_H, for given values of σ_L and σ_H:

$$\frac{\omega_H}{\sigma_L} > \frac{N(2N+1)-(N^2(N-2d+2)+d-1/2)\sigma_L)}{N(N+1)(1-(N-2d+1)\sigma_L)} \tag{2.13}$$

$$\frac{\omega_j}{\sigma_H} < \frac{N(2N+1)+(2N^2-1)d\sigma_L-(N^3+1/2)\sigma_H}{N(N+1)(1+d\sigma_L-(N-d-1)\sigma_H)} = \frac{T_0}{\sigma_H} \tag{2.14}$$

I define a threshold level T_0 for the value of ω_j, which is the maximum level of environmental concern that poor countries can have (for given values of σ_L and σ_H). If ω_j were greater than this value, this would imply that this poor country would be willing to join the d rich country coalition, which is ruled out by assumption.

2.3.1 Grand coalition stability in the four cases

I will verify that each country prefers to join the grand coalition without any profit sharing. A second option, which derives directly from this one, is to assume that the coalition can share the extra profit with all coalition members (condition 2.8). This second condition is just the sum of each country's incentive to join the coalition: $\sum_{j=1}^{N}(\Pi_j^s - \Pi_j^{oo}) > 0$. Since this condition depends on the value of ω_j (of each member), checking for the coalition IS is equivalent to verifying that the average of ω_j meets these same conditions (those to be found).

The objective of this analysis is to check for which parameter set of ω_j, σ_L and σ_H, the different cases ((1) to (4)) work better than others. Since this is an analysis in a volume (restricted by the constraints previously mentioned), I find different conditions for ω_j, for given values of σ_L and σ_H, for all admissible pairs of σ_L and σ_H (and therefore, from now on, I will just talk about the conditions for ω_j, understanding that these are for given σ_L and σ_H). In order to get a better understanding, I will graph these results in the admissible area of (σ_L, σ_H). This area is bounded by: $0 < \sigma_L < \sigma_L^{max}$, $0 < \sigma_H < \frac{1+d\sigma_L}{N}$ (Ineq. (2.10)) and $\sigma_L < \sigma_H$ (the 45° line). In the coming graphs, the area outside the admissible range will be shaded. Finally, we have to remember that the admissible area (σ_L, σ_H) depends also on the values of d and N. For this matter, different situations of d and N will be plotted.

Case (1)

For Case (1), I have no BT and no TT. I call the incentive to join the coalition:

$$^{(1)}\Delta\Pi_j^{N-1} = \Pi_j^s - \Pi_j^{oo} \tag{2.15}$$

where (1) refers to Case 1 (and hence, (2) will refer to Case (2), etc.), j denotes a country with environmental concern ω_j, and $N-1$ refers to this inequality when the country is evaluating whether or not to enter a coalition of size $N-1$. Therefore, Π_j^s is the profit this country has when joining the grand coalition, and Π_j^{oo} is its outside option, meaning the profit when it stays out and the $N-1$ others form a coalition. Using the firm profit, consumer surplus (CS), and environmental damage equations (the first three terms in equation (2.7)), I get:

$$\begin{aligned} ^{(1)}\Delta\Pi_j^{N-1} = &-\frac{N^2\sigma_H}{(N+1)^2}\big(2+2d\sigma_L+(N-2d-2)\sigma_H\big) \\ &-\frac{\sigma_H}{(N+1)^2}\big(N-d\sigma_L-(N-d-1/2)\sigma_H\big)+\frac{\omega_H\cdot N}{(N+1)}\big(1+d\sigma_L+(N-d-1)\sigma_H\big) \end{aligned} \tag{2.16}$$

The first two terms, which are the firms and CS losses, are always negative. In the first case, this is due to the fact that the firm is now facing full abatement, where in the outside option it faced no abatement at all. It is also lowering its production level, since now there is no carbon leakage (from which it was profiting before). In the case of the CS, since the global production goes down (due to the increase in its global costs), consumers also lose. Finally, the last term is always positive, and it is the suppression of the environmental damage. One thing to notice is that this improvement increases with the carbon leakage. This means that incentives to join the GC increase when the carbon leakage increases. Rearranging terms, we have that for the grand coalition to be stable, the following condition must hold:

$$\frac{\omega_j}{\sigma_H} > \frac{N(2N+1)+(2N^2-1)d\sigma_L+(N^2(N-2d-2)-(N-d-1/2))\sigma_H}{N(N+1)(1+d\sigma_L+(N-d-1)\sigma_H)} = \frac{T_1}{\sigma_H} \tag{2.17}$$

In the same fashion as in the previous subsection, I define a new threshold level T_1 for ω_j. This means that if $\omega_j > T_1$, this country will join the GC.

Case (2)

For Case (2), we now have a border tax t and no technology transfer. As before, I find $^{(2)}\Delta\Pi_j^{N-1}$. Before going into the equation, note that for the CS there will be no change with respect to Case (1). Looking at equation (2.3), the CS of a non-signatory does not depend on t. On the other hand, when the grand coalition forms, there is no more tax

imposed. Hence, the tax does not affect the CS, and therefore, there is no difference from Case (1).

Replacing the original equations as in Case (1) and writing ${}^{(2)}\Delta\Pi_j^{N-1}$ as a function of ${}^{(1)}\Delta\Pi_j^{N-1}$, we get the following expression:

$$ {}^{(2)}\Delta\Pi_j^{N-1} = {}^{(1)}\Delta\Pi_j^{N-1} + \frac{tN(N-1)}{(N+1)^2}\Big[2\big(1+d\sigma_L+(N-d-1)\sigma_H\big) - tN - \omega_j(N+1)\Big] \tag{2.18} $$

Looking inside the square brackets, we have a negative term with ω_j. This means that the environmental gain that we found in Case (1) has decreased. This is due to the following: Noting that this gain was bigger the bigger the carbon leakage was, and knowing that when a BT is imposed the carbon leakage decreases, we find that the previous environmental gain decreases as well. On the firm side (which is the other part of the expression inside the square brackets), we have a positive effect if the tax is below some threshold, although not imposing new limits, due to condition (2.12). This effect comes from the reduction of carbon leakage. Therefore, the free-rider in Case (2) is gaining less (from carbon leakage) with respect to what he was doing in Case (1).

Thus we find that Case (2) works for a wider set of parameters $(\omega_j, \sigma_L, \sigma_H)$ if the following condition holds:

$$ tN + \omega_j(N+1) < 2\big(1+d\sigma_L+(N-d-1)\sigma_H\big) \tag{2.19} $$

Noting that Eqn. (2.19) restricts a linear combination of t and ω_j, we can derive the condition in which Case (2) is better than Case (1) by verifying the limiting case when $t = 0$ and obtaining the following expression:

$$ \omega_j < \frac{2\big(1+d\sigma_L+(N-d-1)\sigma_H\big)}{(N+1)} = T_2 \tag{2.20} $$

Hence, if this condition holds, we have that Case (2) can sustain the grand coalition for a wider range of parameters. Again I define a new threshold level T_2 for ω_j (for coming analysis). Knowing these values (σ_L, σ_H and ω_j), we can find an 'optimal' value for the border tax t^* that maximizes the gain for the last country accessing the coalition in Case (2). Maximizing expression (2.18) with respect to t gives us:

$$ t^* = \frac{2\big(1+d\sigma_L+(N-d-1)\sigma_H\big) - \omega_j(N+1)}{2N} \tag{2.21} $$

Case (3)

For Case (3), we have a technology transfer to r recipient countries (meaning that these will have their marginal abatement cost lowered from σ_H to σ_L) and no border tax. As I just did with Case (2), I will represent the IS condition in function of ${}^{(1)}\Delta\Pi_j^{N-1}$. We now get the following expression:

$$ {}^{(3)}\Delta\Pi_j^{N-1} = {}^{(1)}\Delta\Pi_j^{N-1} + \frac{r(\sigma_H - \sigma_L)}{(N+1)^2}\Big((2N^2-1))\sigma_H - N(N+1)\omega_j\Big) \tag{2.22} $$

The intuition of this result goes as follows: Firms' profits will go down (except for the recipients, for whom this is not the case) because of the reduction in σ_S (recall Eqn. (2.5)). The reduction in production level is the same for signatories and non-signatories, but since non-signatories are producing more, they are the ones losing the most (the term is squared in the firm's profit function). Therefore, the impact for the difference in the firms' side is positive. On the consumer surplus side, we have the exact opposite effect. Finally, for the environmental damage, we again get the effect from the previous case due to the reduction of the carbon leakage gained with the decrease of σ_S. These effects are represented inside the last big brackets in the same order stated here. This term is also multiplied by the difference in the marginal abatement costs (hence the technical improvement) and the number of recipient countries r.

As in Case (2) we get again a threshold level for ω_j, which make Case (3) better than Case (1). Equating for the last big bracket, we have:

$$ \omega_j < \frac{\sigma_H(2N^2-1)}{N(N+1)} = T_3 \tag{2.23} $$

Therefore, if this condition holds, more transfers (bigger values of r) make the GC stronger. But these technological transfers have a cost: K per recipient. Therefore, we have to verify if rich countries are willing to pay for these transfers. At this point we have two options: either the GC is not IS in Case (1) or it is. If it is IS, rich countries do not have any incentive to make transfers, unless they want to make the GC more stable. On the other hand, if the GC is not IS, rich countries are willing to pay for these transfers provided that gains from switching to the GC (from the non cooperative equilibrium) exceed the technological transfer costs. I then proceed to calculate the gain for rich countries, which is the difference between profit in the GC and the one with the original small coalition of

d rich countries:

$$\Delta\Pi_d^{d\to N}(r) = \frac{(N-d)\sigma_H - r(\sigma_H-\sigma_L)}{(N+1)^2}\Big(N-(2N(N-d+1))\sigma_L + (N+1/2)(N-d)\sigma_H - r(N+1/2)(\sigma_H-\sigma_L)\Big) + \frac{\omega_H N(N-d)(1+d\sigma_L)}{(N+1)} \quad (2.24)$$

Since this last expression is decreasing in r, rich countries will just transfer to the minimum amount of r recipients in order to make the GC stable – i.e., ${}^{(3)}\Delta\Pi_j^{N-1}(r^*) > 0$. At this point, we have to check if it is profitable for rich countries to do so, since they also have to bear the cost of the technological transfer K. Therefore, r^* will be the minimum natural number for which these two conditions hold:

$$ {}^{(3)}\Delta\Pi_j^{N-1}(r^*) > 0 \quad (2.25)$$

$$ \Delta\Pi_d^{d\to N}(r^*) - (K/d)\cdot r^* > 0 \quad (2.26)$$

If there is no natural number r that make these two conditions hold, then $r^* = 0$, meaning that it is not profitable for the rich countries to transfer technology. Either it does not induce the GC, or it is too expensive compared to the gains attained by switching to the grand coalition.

Case (4)

For this last case, both a border tax t and a technology transfer to r recipients are used. In order to make the analysis simpler, let me define the following expression:

$$ {}^{(i)}\Phi_j^{N-1} = {}^{(i)}\Delta\Pi_j^{N-1} - {}^{(1)}\Delta\Pi_j^{N-1} \quad (2.27)$$

which is no more than the 'gains' of Case (i) with respect to the base case, Case (1). What I mean by 'gain' is that if this value is positive, we have a broader range of parameters where the grand coalition will be stable (I have used this concept in the previous cases without giving it an explicit name). Proceeding as before, we get:

$$ {}^{(4)}\Phi_j^{N-1} = {}^{(2)}\Phi_j^{N-1} + {}^{(3)}\Phi_j^{N-1} - \frac{2r(\sigma_H-\sigma_L)tN(N-1)}{(N+1)^2} \quad (2.28)$$

It is straightforward from this expression that the gains of Case (4) are less than the sum of the gains of Cases (2) and (3). Hence, for Case (4) to be an improvement of either Case (2) and (3), we need the last term to be less than $min({}^{(2)}\Phi_j^{N-1}, {}^{(3)}\Phi_j^{N-1})$. Having any of these gains of Case (2) or (3) (${}^{(2)}\Phi_j^{N-1}$ or ${}^{(3)}\Phi_j^{N-1}$) be negative will imply that Case (4) is never the best solution (concerning an improvement of the solution space for the GC).

For example, if ${}^{(3)}\Phi_j^{N-1}$ is negative, it is straightforward that ${}^{(4)}\Phi_j^{N-1} <^{(2)} \Phi_j^{N-1}$, meaning that Case (2) is better than Case (4).

The intuition for this result goes like this: On one side, having a border tax t reduces the carbon leakage which was the 'driving gain' of Case (3) (when making the transfer). Conversely, the technological transfer r affects the value of σ_S (it decreases it in an amount $r(\sigma_H - \sigma_L)$), which is the driving gain that makes Case (2) better than Case (1) (i.e. ${}^{(2)}\Phi_j^{N-1}$). Therefore, when I use both instruments, I add them, but this combination also diminishes the original gain of each instrument.

Let us assume that both Cases (2) and (3) are better than Case (1), meaning that both ${}^{(2)}\Phi_j^{N-1}$ and ${}^{(3)}\Phi_j^{N-1}$ are positive. Then we could have an improvement moving to Case (4) if the right values for r and t were chosen. Using the condition stated before, we have the following:

$$2tN(N-1) < \underbrace{(2N^2-1)\sigma_H - N(N+1)\omega_j}_{\text{Positive if Case (3) is better than Case(1)}} \tag{2.29}$$

This imposes a new ceiling to the tax t. Furthermore, we should now calculate a new optimal tax value t^{**} that maximizes the expression in Eqn. (2.28). With all this, we should compare the following two possibilities:

$${}^{(2)}\Phi_j^{N-1}(t^*) \lessgtr {}^{(4)}\Phi_j^{N-1}(t^{**}(r)) \tag{2.30}$$

which would tell us which Case, (2) or (4) is the best solution.

The second condition that we have to verify is the following:

$$2r(\sigma_H - \sigma_L) < \underbrace{2\big(1 + d\sigma_L + (N-d-1)\sigma_H\big) - tN - \omega_j(N+1)}_{\text{Positive if Case (2) is better than Case(1)}} \tag{2.31}$$

In the same fashion as before, this condition imposes a ceiling to r, and we should also verify if it is profitable for rich countries to make transfers in this scenario at all (as we did in Case (3)). Assuming that is the case, we should finally compare the following two expressions in order to choose between Case (3) and (4):

$${}^{(3)}\Phi_j^{N-1}(r^*) \lessgtr {}^{(4)}\Phi_j^{N-1}(t^{**}(r^{**})) \tag{2.32}$$

Finally, depending on which solution is preferred in expressions (2.30) and (2.32), we can derive whether case (2), (3), or (4) is best. Therefore, for Case (4) we do not have a specific threshold value, but instead, if $\omega_j < T_2$ and $\omega_j < T_3$ (meaning Cases (2) and (3) do better than Case (1)), then Case (4) could be the best solution, depending on the result

of the analysis just discussed.

Overall analysis

In this subsection I compare different cases and verify which has the biggest gain for the last signatory to join. This could induce the GC when it was not present, or it could strengthen the GC.

In order to analyse these cases, we should examine all different instances for ω_j – i.e., when it is greater or smaller than combinations of these thresholds. This would imply checking 4! possible orderings. Fortunately, we can note that: $T_1 < T_0$, $T_1 < T_2$, $T_3 < T_0$ and $T_3 < T_2$.[(14)] We also have that $T_1 < T_3$ iff $\sigma_H > (N+1)/(N^3 + 1/2)$. Finally, $T_2 \lesseqgtr T_0$ depending on a more complex condition that varies with all parameters. Using these four inequalities, we have that $\{T_1, T_3\} < \{T_0, T_2\}$, which yields to four possible orderings of the thresholds. However, if $T_3 < T_1$ we have that $\sigma_H < (N+1)/(N^3 + 1/2)$, which in turn implies that $T_0 < T_2$, and therefore, we end up with only three cases:

A: $T_3 < T_1 < T_0 < T_2$

B: $T_1 < T_3 < T_0 < T_2$

C: $T_1 < T_3 < T_2 < T_0$

[(14)]These proofs are available upon request. For the first case, I assume that $d < N/2$ and $N > 3$. For the second and third cases I suppose that $2(d+1) < N$. All these assumptions are only sufficient. No condition is need for the fourth inequality.

In other words, depending on the values of σ_L and σ_H (and also d and N), we have these four orderings. This means that we can depict these cases in different areas in the (σ_L, σ_H) plane. Consequently, I use a graphical representation of these cases, depending on the values of σ_L and σ_H (for given values of d and N):

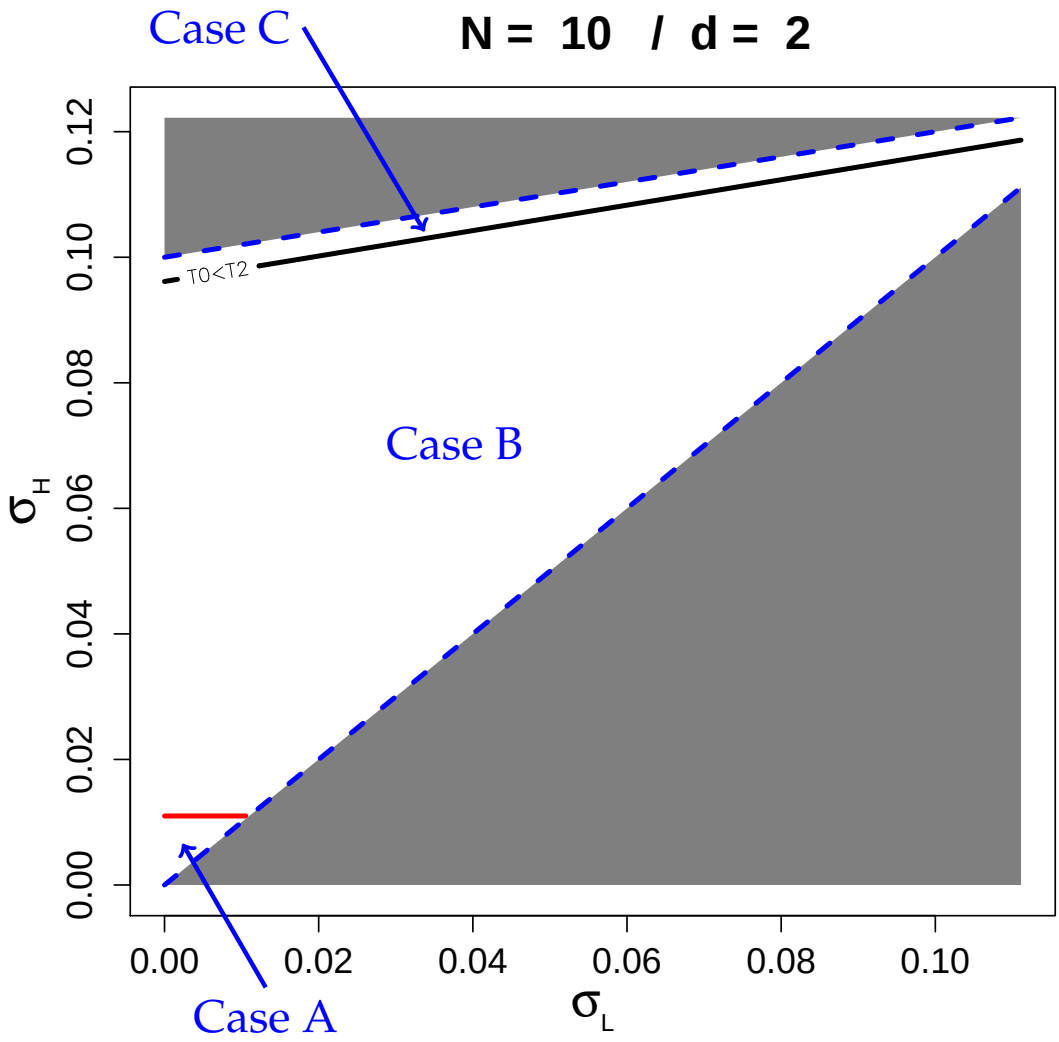
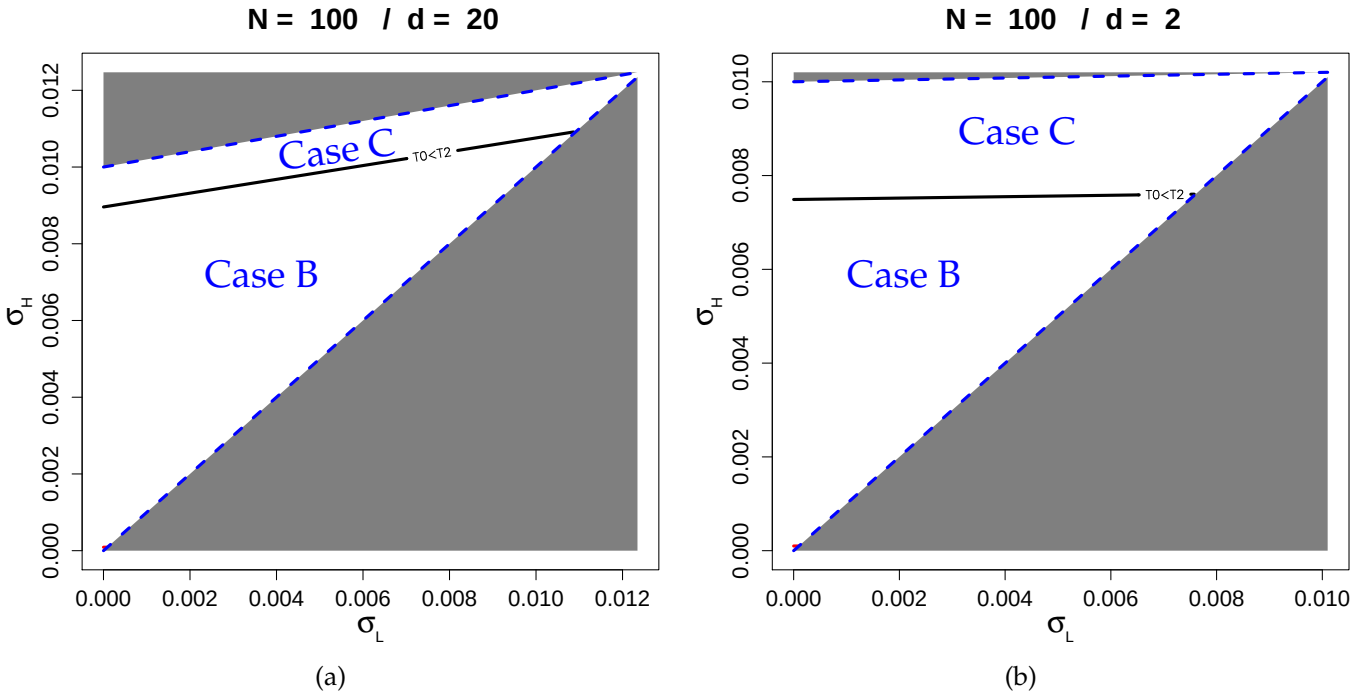


Figure 2.1: Graphical representation of thresholds' cases.

The dashed lines represent the limits of the admissible values for any pair (σ_L, σ_H). The 45° line simply corresponds to $\sigma_L < \sigma_H$. The upper dashed line comes from condition (2.10). For clarity, the inadmissible area has been greyed. Since $T_1 \lesseqgtr T_3$ only depends on σ_H (and not of σ_L), we get a horizontal line (dividing Cases A and B), which is always below of the $T_0 \lesseqgtr T_2$ line. Furthermore, as N increases the former line goes further south,[(15)] becoming almost indistinguishable from the corner, as can be noted in the fol-

[(15)]This comes directly from the result that $T_1 < T_3$ iff $\sigma_H > (N+1)/(N^3 + 1/2)$. Hence, if N grows, the

lowing two examples:



(a) (b)

Figure 2.2: Two more examples of thresholds' cases.

We can also observe that lower values of d (fewer rich countries) make the upper area (Case C) bigger. Now let us focus on the possibilities of ω_j. Following the same order of these three cases, we have (I suggest reading comments from right to left):

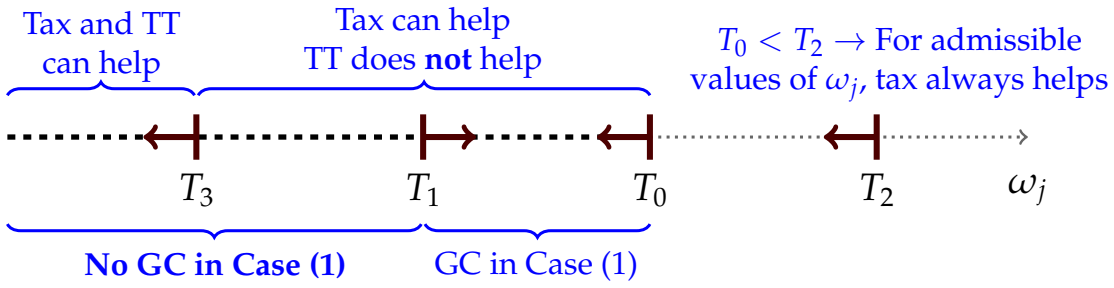


Figure 2.3: Overall analysis (Case A)

Let us recall that $\omega_j < T_0$ holds by assumption (Inequality (2.14)). To make this point clear graphically, the inadmissible area of ω_j has been represented with a grey dotted line

RHS terms tends to zero rapidly.

(instead of the black dashed line). Observing Cases A and B, when $T_0 \leq T_2$, we see that it is always better to use a border tax in order to expand the range of parameters that sustains the GC (or to make the GC stronger).

Continuing only with case A, where $T_1 < \omega_j < T_0$, we are in the case where the GC exists in Case (1) and, as noted before, a tax can make it stronger. Here, transferring technology does not improve the situation. If $T_3 < \omega_j < T_1$, we find ourselves in the case where there is no stable GC in Case (1) and only the border tax can induce it. Finally, when $\omega_j < T_3$, transferring technology becomes an option (assuming that $r^* > 0$), and the final recipe (BT, TT, or both) will depend on the result of conditions (2.30) and (2.32).

Following the same reasoning for case B, we have:

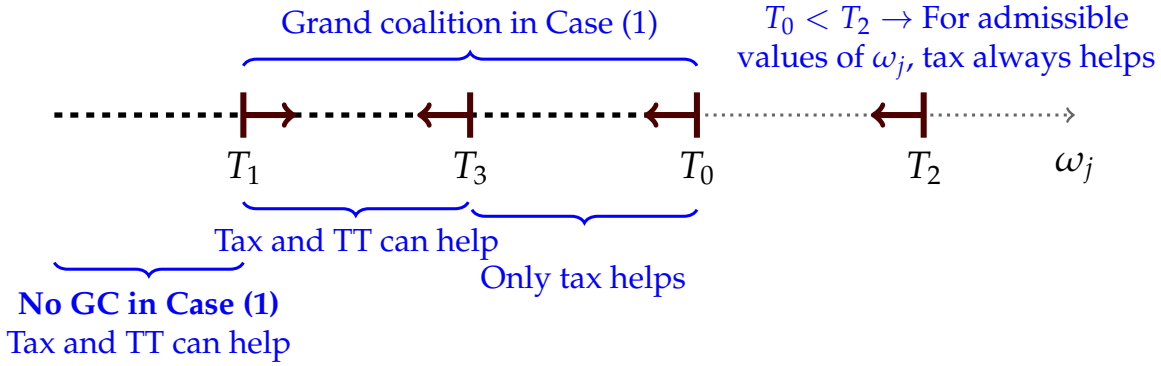


Figure 2.4: Overall analysis (Case B)

Case B is very similar to case A. The only difference is that we have inverted the order of T_1 and T_3. This translates into adding a range for ω_j where the GC is stable in Case (1) ($T_1 < \omega_j < T_3$) and both a tax and a technological transfer can strengthen it. If $T_3 < \omega_j < T_0$, only the tax can make the GC stronger. Finally, when $\omega_j < T_1$, there is no GC in Case (1), and both instruments can induce it. In third place we have case C, with:

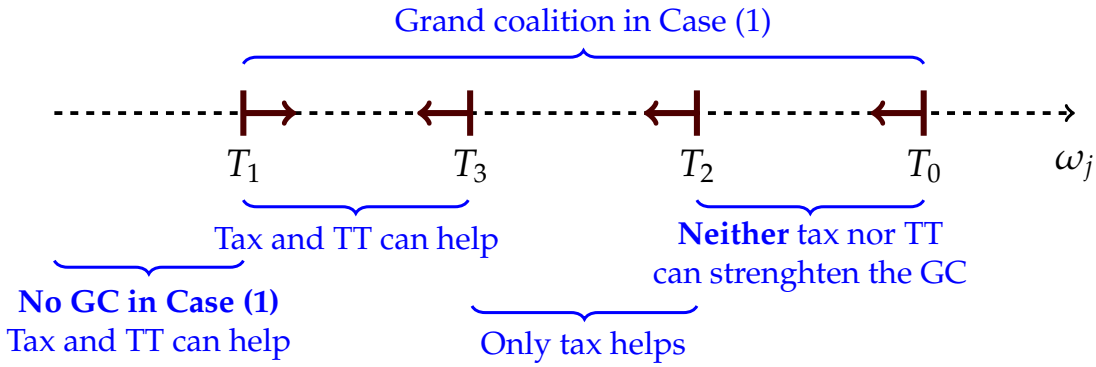


Figure 2.5: Overall analysis (Case C)

The only difference between case B and C is that in C there is a range of $T_2 < \omega_j < T_0$ where the GC is stable, but neither instrument can improve its stability (this is new compared to Cases A and B). As stated after Figure 2.2, Case C becomes more relevant when N is big and d is small (Figure 2.2b). Therefore, when σ_H is close to its upper limit, we are in the presence of this case. In general, though, it seems that Case B is the predominant one.

It is interesting to note that since $T_3 < T_2$, we know that a border tax works in more cases than a technological transfer. In other words, there are more parameter values where the tax works and the transfer does not. When the transfer can work (depending if $r^* > 0$), a tax would work too, the aforementioned Case (4). It is worth recalling that depending on the results of conditions (2.30) and (2.32), it could be the case that the tax is not preferred and the transfer still is.

If the GC is not stable in Case (1), we usually are in the situation where both instruments can work (Case (4)). The only case where this is not true is Case A, when $T_3 < \omega_j < T_1$ and only the border tax helps, but recalling Figures 2.1 and 2.2, it seems that Case A is a rare one. The good news is that since $T_1 < T_2$ (as it always is), we know that if there is no GC in Case (1) (i.e. $\omega_j < T_1$), the tax always does better, since $\omega_j < T_2$. The only case where neither instrument can do any better is when $T_2 < \omega_j < T_0$ in Case C, but here we are already in presence of the GC in Case (1).

Up to this point, I have done the analysis assuming countries inside the coalition are not sharing their profits. In other words, for the GC to be stable we need the condition that no country wants to leave, which translates into having the least-green poor country wanting to stay in the GC. This is the value of ω_j discussed so far, which actually is $\min_{j \notin \mathcal{D}, j \notin \mathcal{R}}(\omega_j)$, the least-green poor country that has not received a TT. A second option that can be studied is one where coalition members (in this case, the GC) share their profits. This means that inequality (2.8) holds. Recalling the country profit formulation, we can perform the whole analysis done for far by replacing ω_j with the mean of ω_j just mentioned, getting the same results.

At this point it is worth noting that asymmetry of poor countries can play an important role. If poor countries are symmetric, the no-sharing solution suffices for the general result. On the other hand, if countries are asymmetric, the possibility of sharing profits by signatories changes the result. In the no-sharing case, we need to verify for the least-green non-recipient: It resembles the weakest link problem. But in the sharing case, which is more profitable for the coalition and therefore more interesting, things change.

To see how, let us compare two cases: one with symmetric poor countries and one with asymmetric ones, both having the same average ω_j. Therefore, when we are in the presence of TT, non-recipients' environmental concern average $((\sum_{j\notin\mathcal{D},j\notin\mathcal{R}}\omega_j)/(N-d-r))$ will change depending on the recipients chosen. This in an improvement with respect to the symmetric case (since in that case we always get the only value of ω_j) and, as it will be proven in Section 2.4, the optimal option is to transfer to the least-green poor countries.

2.3.2 Retaliation tax

So far I have considered that if signatories tax imports coming from non-signatories, the latter will not retaliate with another border tax. In this section I will relax this assumption to examine the incentives of a non-signatory to retaliate. It is good to recall from Brander et al. (1984) that in a set-up of imperfect competition (as this one), countries have the private incentive to impose import tariffs, but they jointly gain more if they enter into a free trade regime. In other words, they face a prisoner's dilemma problem that can be solved with coordination. Therefore, up to this point, in the present set-up, countries have coordinated into free trade.

However, if a country or group of countries deviates from this trade equilibrium and imposes a border tax (even for good motives such as environmental protection), its/their counterpart could retaliate with a similar tax. This is the case that I analyse and, as before, I assume that the coalition will impose a border tax t on products coming from non-signatories and that non-signatories will decide separately whether they will impose a retaliation tax t_2 or not. This means that non-signatories act as singletons and not in a coordinated manner.

Following this idea, a non-signatory country will maximize its own profit with respect to the retaliation tax $t_2 \in [0,t]$. Since this program is solved independently by each non-signatory, which is part of a heterogeneous group, I rename the retaliation tax t_2^j. Recalculating the firm profit, consumer surplus, damages from emissions, and the new tax revenue, and only considering those terms affected by t_2^j, we get:

$$\hat{\Pi}_j(t_2^j) = \frac{(1+\sigma_{\mathcal{S}}+kt_2^j)^2}{(N+1)^2} + \frac{(N-\sigma_{\mathcal{S}}-kt_2^j)^2}{2(N+1)^2} - \frac{\omega_j k(N-k)t_2^j}{N+1} + \frac{\left(k-(N-k+1)\sigma_{\mathcal{S}}-k(N-k+1)t_2^j\right)t_2^j}{N+1} \quad (2.33)$$

where $\hat{\Pi}_j(t_2^j)$ is the part of the non-signatory profit that depends on t_2^j. Note that this expression is independent of t, the border tax chosen by signatories. This is in line with

the results of Brander et al. (1984), where optimal taxes of a home and foreign country are independent (they use a two-country model). Verifying the second derivative of this expression with respect to t_2^j, we get that it is always negative. Therefore, making the first derivative equal to zero will give us the optimal tax that a non-signatory j would impose. Doing so, we get:

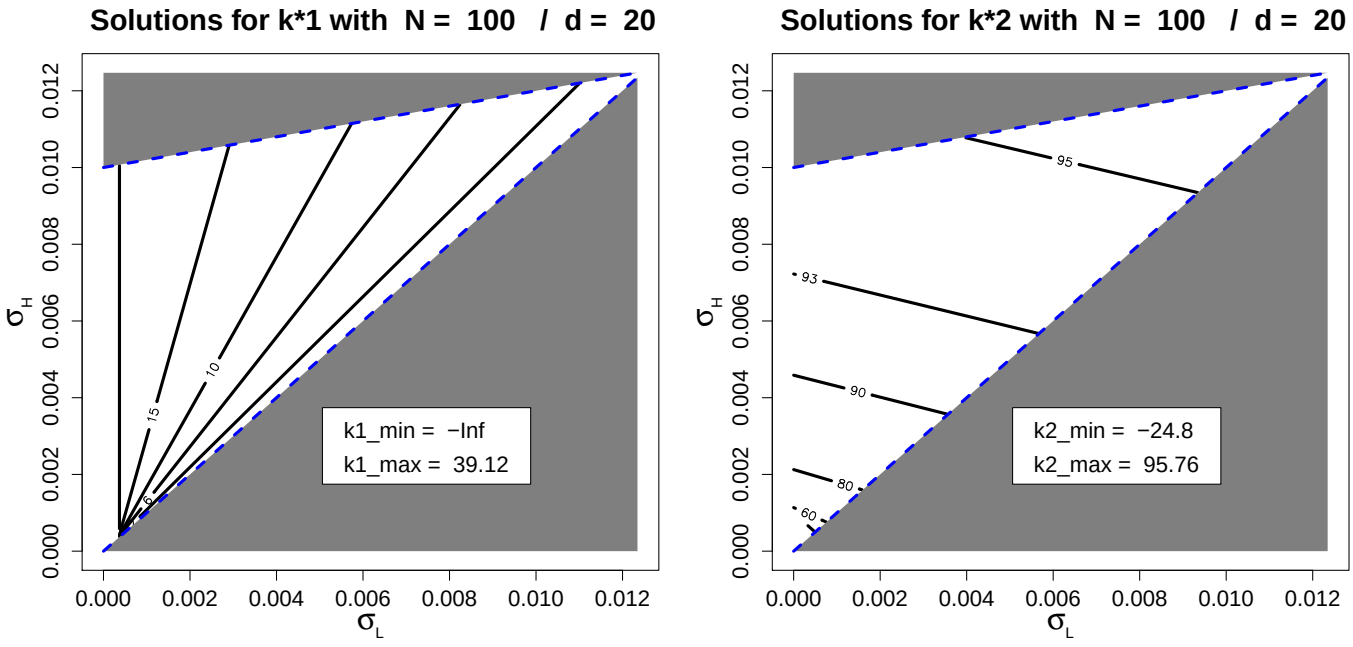
$$t_2^{j*} = \frac{-\Big(3k + (N+1)(N-k+1))\sigma_{\mathcal{S}} - \omega_j(N-k)(N+1)k\Big)}{(3k - 2(N+1)(N-k+1))k} \tag{2.34}$$

Due to the complexity of this expression I will perform two simple analyses. In both I verify when this result is less than or equal to zero, meaning that the non-signatory prefers not to retaliate. First, I check when retaliation is not preferred in the (σ_L, σ_H) plane, as in Figure 2.1. Note that this expression also depends on the environmental concern ω_j and on the coalition size k (previous analyses were made only for the grand coalition). Realizing that the environmental concern only *helps* to make the retaliation tax equal to zero, I will assume the worse case, meaning that $\omega_j = 0$, and check which values of k this inequality holds for. Since we know that the denominator is negative (it is the second derivative of the original program), we find that this condition translates into the following:

$$k(3 + (N+4)\sigma_{\mathcal{S}}) \leq (N+1)^2\sigma_{\mathcal{S}} \qquad \text{with} \qquad \sigma_{\mathcal{S}} = d\sigma_L + (k-d)\sigma_H \tag{2.35}$$

Solving this quadratic inequality, we finally get k^{*1} and k^{*2}. This means that for $k \in [k^{*1}, k^{*2}]$ the previous inequality holds, and therefore, non-signatories choose not to retaliate.

This area enlarges if $\omega_j > 0$. Since these two expressions are complex again, I plot them in the (σ_L, σ_H) plane, getting:



(a) (b)

Figure 2.6: Contours for k^{*1} and k^{*2}

Figure 2.6a shows the contours for different solutions of k^{*1}. Observing that plot, we realise that these values decrease when the lines 'rotate' clockwise. Take for example $k^{*1} = 15$. For all points to the right of this line – i.e., the pairs (σ_L, σ_H) – we have that $k^{*1} \leq 15$. This means that for coalition of size $k \geq 15$ we know that $k \geq k^{*1}$. In other words, this inequality holds for points in this region. The same reasoning can be performed for k^{*2} in Figure 2.6b, and we get a second area that for a given k, $k \leq k^{*2}$. In this case, the area begins in one of these contour lines and expands to the northeast. Hence, for a particular coalition size k, we can intersect these two regions, and we get the pairs (σ_L, σ_H), for which $k \in [k^{*1}, k^{*2}]$. In other words, for that region and coalition size k (meaning the triple (k, σ_L, σ_H)), the non-signatory prefers not to retaliate. Note that this result improves if the non-signatory has positive environmental concern (meaning that non-signatories do not retaliate in a larger area). Therefore, for a considerable amount of (σ_L, σ_H) cases, small and medium size coalitions are not retaliated against.

The second analysis focuses on when the grand coalition forms; we want to see what happens if a country leaves the coalition. Would it be in its interest to retaliate? Using the solution in Eqn. (2.34) and knowing that the denominator is always negative, we can again find a threshold for ω_j that tell us when this country is not willing to retaliate. This

condition and the according graph in (σ_L, σ_H) are:

$$\omega_j \geq \frac{3(N-1)+(N-5)(d\sigma_L+(N-d-1)\sigma_H)}{N^2-1} = T_4 \tag{2.36}$$

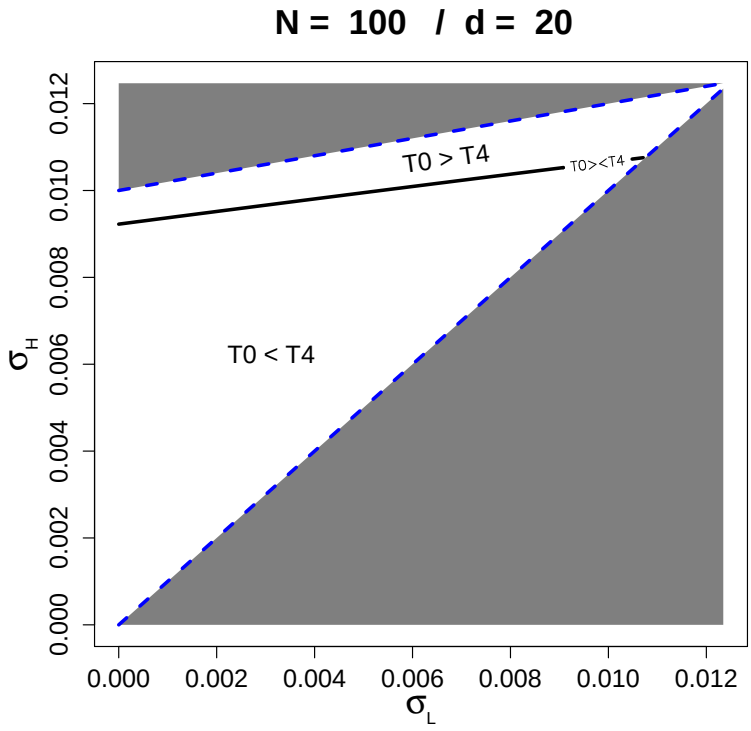


Figure 2.7: Graphical representation of thresholds T_4.

As we can observe here, this is bad news for the grand coalition. $T_0 < T_4$ means that for that area the non-signatory always prefers to impose a retaliation tax (since $\omega_j < T_0$). Moreover, we also get that $T_1 < T_4$ and $T_3 < T_4$ (for any σ_L, σ_H). The first inequality implies that for those cases where there is no GC in Case(1), and we want to use a border tax to induce it (which now we know always works), non-signatories will prefer to retaliate. In other words, when they do not retaliate ($\omega_j > T_4$) we are always in the presence of the GC in Case (1) ($\omega_j > T_1$). The second inequality is less interesting: It simply means that considering that retaliating is profitable for non-signatories, a TT may or may not help depending on the value of ω_j. For T_2 and T_4, there is no rule: $T_2 \lesseqgtr T_4$. This means that there are cases for ω_j, σ_L and σ_H where a BT and a retaliation tax are optimal, only one of these, or none; all these done by the $(N-1)$ coalition and the non-signatory, respectively.

2.4 Recipients of technology transfers

In the first part of this chapter I analyse which country or group of countries should receive a technological transfer if that option would be the optimal situation (hence, we are considering Case (3) or (4)). In the second part, I illustrate the results of using only a border tax or only a technological transfer when compared with the use of both (assuming were are in a situation of Case (4)). The idea here is to give a better understanding of the effects of each instrument and how they work together.

I start by focusing on the country accession. Let us assume that we are in a condition where no country wants to join nor leave: The coalition is stable. If no other country is willing to join the coalition, we consequently know that the greenest outsider is not willing to join. This is simply due to the fact that all outsiders share the same abatement technology σ_H and only differ on their environmental concern ω_j: They can be ordered from the greenest one to the least-green one. Hence, if the greenest one is not willing to join, it is evident from the country profit equation (2.7) that no other country is. It is worth noticing that this is a one-shot set-up, and therefore when I talk about a *sequence* of countries, implying therefore an order, in reality what I am simulating is the thinking process of each country as it decides whether or not it will join a given coalition. We can understand the simultaneous decision process as following: Given a coalition of k countries, the greenest outsider evaluates whether or not it is profitable to join, whatever the other countries do. Obviously, all other countries also think about accessing the coalition at the same time and know that it is profitable for the greenest outsider to join. In that case, it may be profitable for the second-greenest country to join the coalition, no matter what the rest do. In this way, we could have a 'cascading' process ending with all countries joining the coalition. Of course, when I use the word 'cascading', I am not implying a dynamic process, only the strategic reasoning presented. In the same manner, we can clearly see that if this ordered sequence of countries does not produce the 'full cascading' (leading to the grand coalition), no other ordering will. Since all countries know this, they proceed accordingly.[16]

I return to the discussion of the previous section and study the case where both a border tax is implemented and transfers to one or more recipients are possible. The assumption that the BT is implemented is made without a loss of generality (the proof is the same one). No retaliation tax is assumed, though.[17] In other words, Case (4) works the

[16]It could also be considered a dynamic case where countries join sequentially. If this where the case, a discount factor for future gains or losses should be introduced alongside a timing system, which would add more complexity to the model. Since the idea is to keep to model as simple as possible, I do not consider this option, although the reader can visualise this scenario too.

[17]This is assumed for simplicity. Having the possibility of retaliation in the general case escapes the

best. The consequent question that arises is: Which countries should be the recipients? I show that in the presence of a border tax, if a group of $d + r$ (d rich countries and r recipients) produces the cascading, then the set of groups that satisfies this condition always includes the group of outsiders, or poor ones, that are the *least green*. This result can appear counter-intuitive at the beginning, but it should be noted first, that since countries are of the same size, the cost to switch their abatement technology from σ_H to σ_L is always K, regardless of how green the country is. Second, knowing that r countries are recipients and d countries are donors, then $(N-d-r)$ countries are non-recipients, and the goal is to make them produce the cascading. Leaving the greenest countries in the non-recipients group is the best thing to do, since they are more prone to access any given coalition. That means that the countries joining the expanded coalition (donors plus recipients) to form the grand-coalition are the ones that bear the higher abatement costs.

To prove this, I assume that there is a group of r recipients that produce the whole cascading (on top of the d rich countries). Abusing the notation a bit, r will refer both to the amount of recipient countries and to the group of such countries. I show then that if a new recipient group r' is formed, changing one country of the original r group with one less-green country from the non-recipients, then $(d+r')$ also produces the cascading. Iterating the modification of the recipient group, I arrive at $(d+r^*)$, where r^* is the group of least-green outsiders (of the same size of r), which again generates the cascading.

Let us call *Cascading 1* the case where it is supposed that the expanded coalition $(d+r)$ produces the cascading. This means that for each non-recipient i that enters the coalition (in a greenest – less-green 'order' as stated before), the IS condition in (2.8) holds. In the same manner, let us denote *Cascading 2* the case where we have substituted one country from the r group with one country from the non-recipients (this new country being, of course, less green than the replaced one), naming this new recipient group r'. I show that the new expanded coalition $(d + r')$ also produces the cascading, hence:

Proposition 1 *Within the game set-up described above, and with d being the amount of initial rich countries in the coalition and r being a recipient group (of r countries) that produces the whole cascading (Cascading 1) for all i between 1 and $N-d-r$, then the whole cascading is also produced starting from the coalition $(d + r')$, where r' is a less-green recipient group of r countries (Cascading 2):*

$$\underbrace{\Pi_s^{d+r+i} > \Pi_n^{d+r+i-1}}_{\textit{Cascading 1}} \quad \Rightarrow \quad \underbrace{{}'\Pi_s^{d+r+i} > {}'\Pi_n^{d+r+i-1}}_{\textit{Cascading 2}} \qquad \forall i \in \{1, \ldots, N-d-r\} \quad (2.37)$$

where Π_s^{d+r+i} is the non-recipient profit in a coalition of members $(d+r+i)$ (d donors, r re-

scope of this paper and it is addressed as a possible extension.

cipients and i non-recipients), and $\Pi_n^{d+r+i-1}$ is the non-signatory profit of a coalition with one member less (the outside option for the non-recipient i). The prime in $'\Pi$ indicates that we are in Cascading 2, meaning that the recipient group is r' and that the sequence i of non-recipients coming into the coalition has been replaced accordingly. The following diagram shows the non-recipients i sequence for Cascadings 1 and 2:

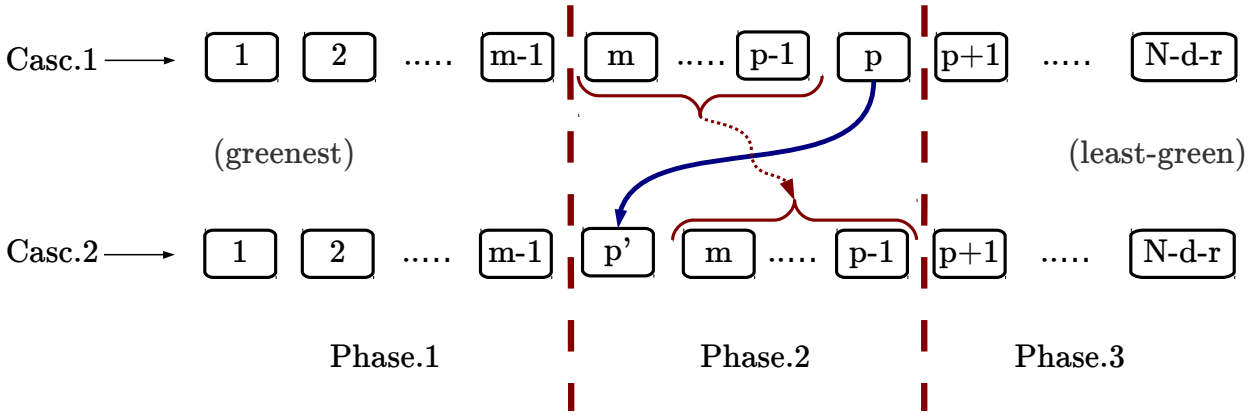


Figure 2.8: Non-recipients sequence for Cascadings 1 and 2.

As noted in Fig. 2.8, the non-recipient sequence has been divided into 3 phases. This comes from the fact that we have replaced one country in r, namely [p], which has entered into r', replacing the country $[p']$ that is now in the non-recipient sequence. Note that when a recipient group becomes less green, the corresponding non-recipient group becomes greener, where $[p']$ is greener than $[p]$. Due to this, the sequence has been modified, where phases 1 and 3 are unchanged and the modification only applies to the countries in phase 2. As stated before, $[p']$ is greener than $[p]$ and therefore enters first in the cascading process, as shown in the previous figure.

The proof of the statement in (2.37) consists on showing that for each phase, the following Inequality 1 and 2 hold:

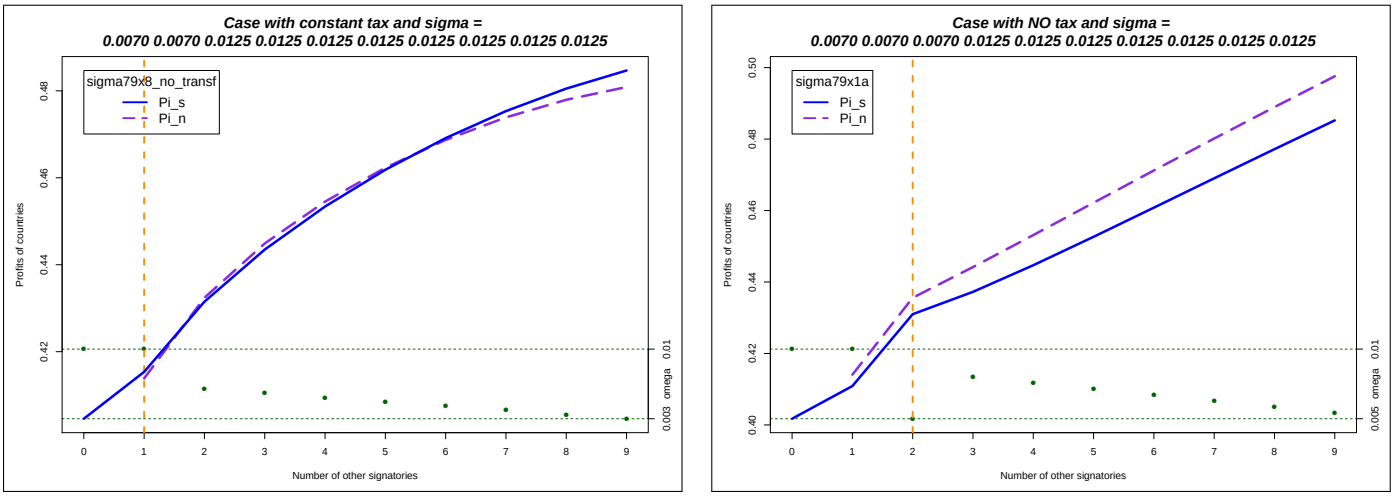
$$\underbrace{\underbrace{{}'\Pi_s^{d+r+i} \geq \Pi_s^{d+r+i}}_{\text{Inequality 1}} > \underbrace{\Pi_n^{d+r+i-1} \geq {}'\Pi_n^{d+r+i-1}}_{\text{Inequality 2}}}_{\text{Cascading 2}} \qquad \overbrace{\qquad\qquad}^{\text{Cascading 1}} \tag{2.38}$$

A detailed proof can be found in Appendix 2.B. This result states that the set of recipient groups that can produce the cascading always contains the group of the r least-green countries within the outsiders.[18]

[18] Depending on the parameters chosen, this set can contain only one group, r^*, or more than one, but it

The idea behind this proof is the following: by hypothesis we know that Cascading 1 holds. Inequality 1 means that the profit of a country joining the modified coalition (denoted by the prime) is at least as good as the one of a country joining the original coalition. This holds for the first country joining ($i = 1$) and thereafter ($i > 1$), the cascading idea. In both cases the comparison is made using the accession order already mentioned (greener countries first). For Inequality 2 we have the non-signatory counterpart. This just means that the outside option of this country, with the new recipient group r', is worse than or equal to the original case. In other words, with the new recipient group r', joining countries do better (compared to the original situation with r) and their outside options worsen (also compared with the same situation).

In order to illustrate the results of the previous section, I give an example where Case (4) is the best one. I start by showing what happens if one instrument at a time is used, and then when both are used, in the most 'efficient' way (transferring to the least-green countries). To do so, I simulate using 10 countries, in which the initial coalition is composed of the two rich countries. To visualize the decision-making process I use the same type of graphic representation as in Barrett (1997), Carraro (1999), and Diamantoudi and Sartzetakis (2006).[19]



(a) Border tax, no transfer.

(b) Transfer, no border tax.

Figure 2.9: Base case.

Fig. 2.9a illustrates the case where the border tax is implemented without technologi-

always includes r^*.

[19]Graphs used in the cited papers considered symmetric countries. Since the goal is to analyse IS (condition 2.8), dividing this inequality by k gives us a *normalized* version of this inequality (on a *per country* basis). Hence, we get: $(\sum_{j\in\mathcal{S}_k} \Pi_j^s)/k > (\sum_{j\in\mathcal{S}_k} \Pi_j^{oo})/k$. The plotted lines are each side of this inequality.

cal transfers. The border tax rate is assumed to be equal to the marginal cost of abatement outsiders would have incurred if they had abated (σ_H). Fig. 2.9b presents the reverse case, where one country benefits from the technological transfer ($r = 1$) and no border tax is implemented. Both cases are modelled using the following parameters: $\sigma_L = 0.0070$, $t = \sigma_H = 0.0125$, $\omega_H = 0.100$, $\omega_L = 0.060 \ldots 0.030$.[20]

In both cases, the solid line represents the mean coalition profit and the dashed line the mean outside option. The horizontal axis presents the number of countries in the coalition. Since the countries are heterogeneous, the accession ordering is important. Therefore, to depict this sequence I introduce round dots; these account for ω_j of each country, represented in the secondary y-axis. The d rich countries (equal to two in this example) are the first ones in the graph, on the left. In Fig. 2.9a no country receives any transfer, and following the previous reasoning (page 84), I only need to test if the greenest – least-green ordering suffices, which is the one depicted. When a transfer is considered, the question lies in the identity of the recipient(s). I plot one option in which the least-green country is the one receiving the transfer, signified by the third dot in Fig. 2.9b. The rest of the sequence is identical to the one used in Fig. 2.9a. Choosing another recipient does not change the result. Finally, the value of σ_j for each country (following their order) is printed on the top of the graph, and the vertical dashed line shows the transition from countries with good technology (the expanded coalition) to those without it (non-recipients).

We observe that in the first case, implementing a border tax can alleviate the burden carried by the rich countries by reducing carbon leakage, increasing coalition firms' profits, and getting extra revenues (from taxes), but it needs the use of an MPC in order to trigger the grand coalition formation. The result in the second case is worse, since the absence of the border tax makes the rich countries much worse off and the unique technology transfer does not induce the recipient country to join (and stay) in the coalition.

[20] These values were chosen in order to: a) resemble the examples shown in Barrett (1997), b) have all countries producing a positive amount of good and, c) have a small coalition in the base case.

Let us observe an example of a transfer that induces the GC without the need of an MPC in the following figure.

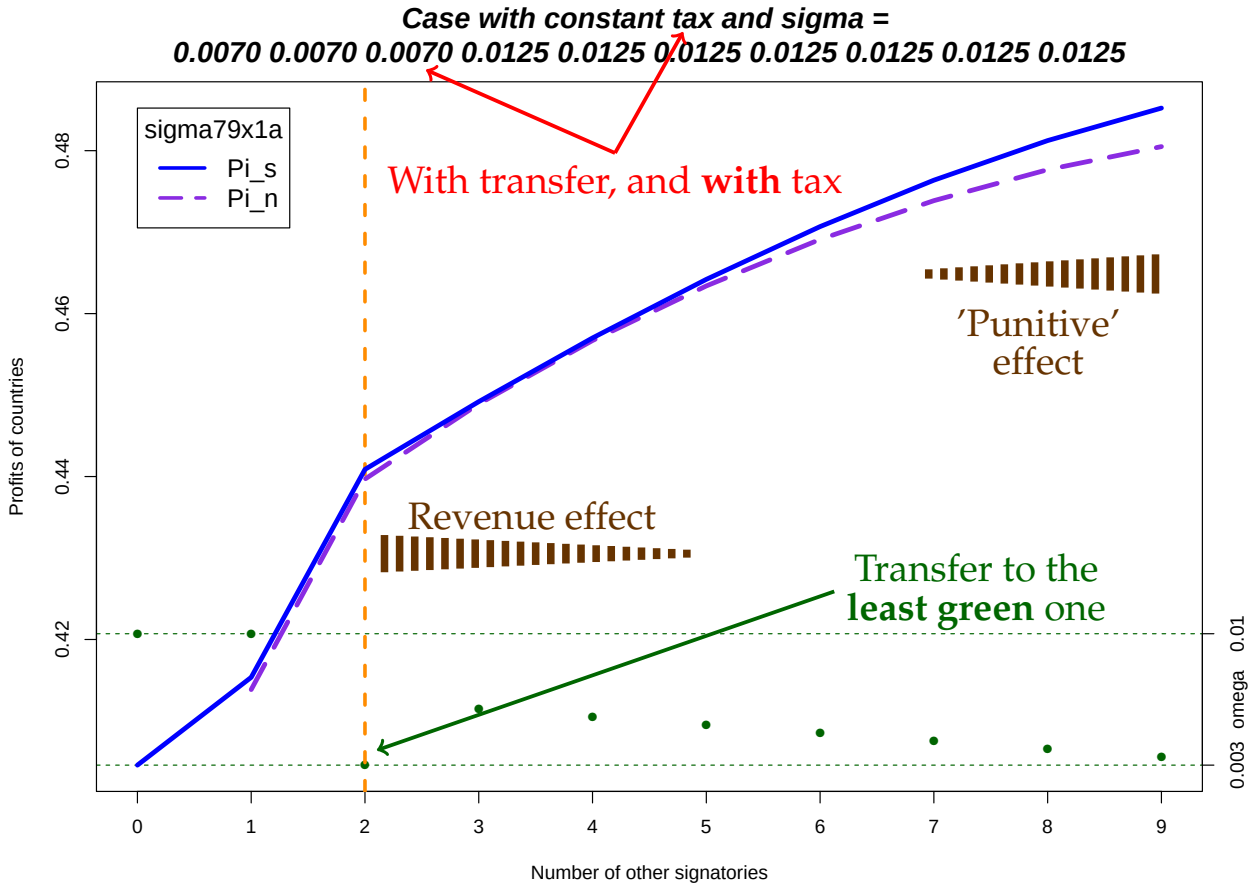


Figure 2.10: After Tax and Transfers.

I have used the same parameters as in Fig. 2.9a and Fig. 2.9b, and the GC without a MPC has been triggered with only one recipient, the least-green outsider. The intuition is as follows: The technology transfer 'buys in' the least-green country, putting it inside the expanded coalition and changing its incentive for abating. The initial coalition is enlarged, and the border tax helps to sustain it. The cascading then occurs, starting with $(d+r)$ countries. In contrast to what happened in Fig. 2.9a without any transfer, here the IS condition continues to hold even if the revenue effect coming from the border tax diminishes as countries join the coalition. When more countries join the coalition, the border tax has a punitive effect, in the sense that non-signatories' exports are facing a disadvantage with respect to the rest of the world. Nevertheless, after the reduction of the revenue effect and before the apparition of the punitive effect of the border tax, there is a critical period where the IS condition might not hold. This cascading process mimics a mountain crossing[(21)] (in this case peaking around $k = 3$ or 4), where the outside option is the

[(21)] As in the Mountain Crossing theorem.

mountain to be crossed and the mean coalition profit is the maximum altitude we reach at each step. With only the border tax, the domino effect stops at some point: The mountain cannot be crossed. But combining the border tax *and* a transfer (when leaving out the greenest countries) allows us to make it through. Both instruments reinforce themselves.

2.4.1 Disposing of the MPC

Let us assume now that we want to induce the grand coalition without the use of an MPC. This could be due to different reasons – for example, an acceleration of the starting date of meaningful abatement (since we do not need to wait for the MPC amount of countries to ratify the agreement). Therefore, having determined that if the rich countries want to make a technology transfer in order to induce the grand coalition, they have to start with the least-green countries, let us now address the question of how many r^* recipients we need in order to produce the full cascading.[22] Unfortunately, equations developed from the IS condition get much too complex to help create a readable analytical solution for this question. I therefore rely on numerical simulations. I define a $\Delta\Pi^s_{oo}(d,r,i)$ function, which is just inequality (2.8) with all terms put on the left side. Therefore, the IS condition holds if the function is positive. Hence, we have:

$$\Delta\Pi^s_{oo}(d,r,i) = \sum_{j\in\mathcal{S}_k} \Pi^s_j \; - \sum_{j\in\mathcal{S}_k} \Pi^{oo}_j \tag{2.39}$$

Therefore, if for a given value of r, $\Delta\Pi^s_{oo}(d,r,i)$ is strictly positive for all possible i's (d is given), then we get the full cascading. Defining this function makes use of the fact that outsider countries that should be targeted to receive a technology transfer are perfectly known from proposition 1 for each level of r. In Fig. 2.11 we can observe a continuous version of this function[23] using the same parameters as in the previous examples.

The figure shows the solution area where $\Delta\Pi^s_{oo}(d,r,i) > 0$, and more importantly, its limit $\Delta\Pi^s_{oo}(d,r,i) = 0$. Therefore, it is easy to see that the solution for this case is $r^* = 1$. With $r = 0$ we have the base case where there is no transfer and the border tax is implemented. Following the vertical dashed line at $r = 0$, we can observe that we have the same result as before. For values of i between 0 and 0.8 and then from 3.2 until the end, the value of $\Delta\Pi^s_{oo} > 0$. In the range left in between, it is negative, meaning that for this case we are in need of an MPC if we want to reach the grand coalition (we are not able to cross the mountain). In the case of $r = 1$, we can clearly see that $\Delta\Pi^s_{oo} > 0$ for the full range of i, meaning that full cascading occurs. Following this reasoning, r^* can be found by check-

[22] Of course, it is assumed that this r^* is a profitable solution, as discussed in subsection 2.3.1, page 69.

[23] This is coming directly from the country profit function. The only 'trick' was that I had to create a continuous version of $\sum\omega_j$ to be used in the damage part of this function. The domain of $\Delta\Pi^s_{oo}$ is restricted to octant I (+++) and with $(d+r+i) \leq N$, which is the area of interest.

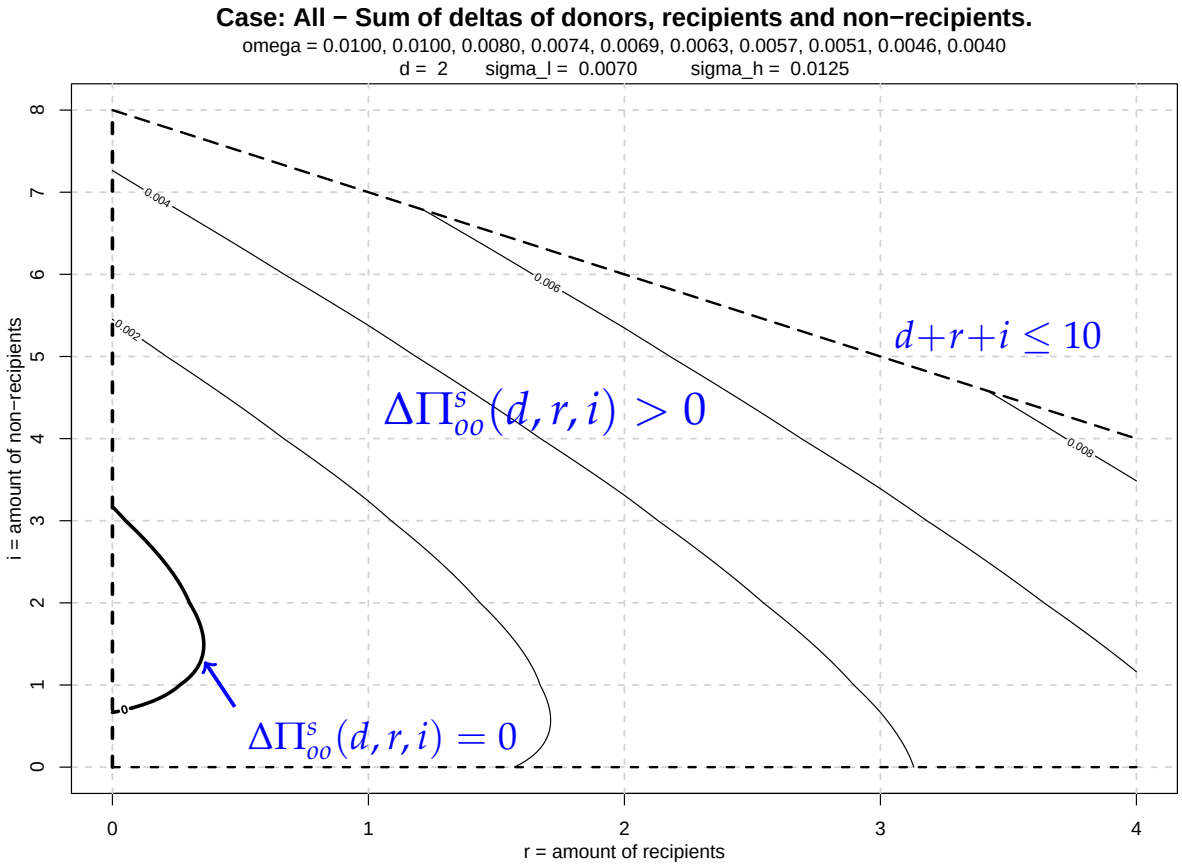
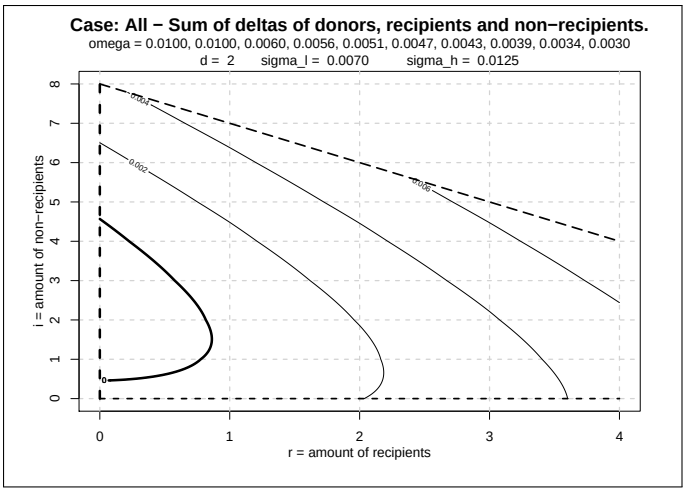


Figure 2.11: Contour of function $\Delta\Pi^s_{oo}(d,r,i)$, with $d = 2$.

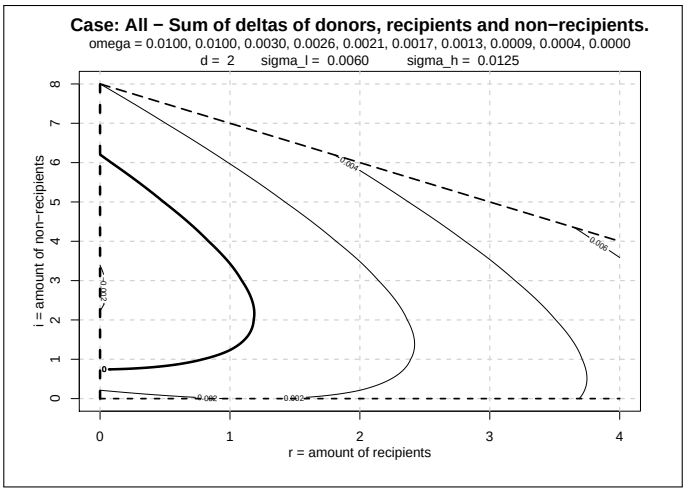
ing when $\partial r/\partial i = 0$ in the implicit function between r and i, given by $\Delta\Pi^s_{oo} = 0$. Hence, we find a maximum $\hat{r}$ (of r with respect to i in this implicit function), and r^* is just the integer solution for $r^* > \hat{r}$. Actually, $\hat{r}$ can cross some integer, say r_1, and r^* can still be equal to $r_1 - 1$. This is due to the fact that we only have to check that, for a given integer level r, $\Delta\Pi^s_{oo}(d,r,i) > 0$ for all possible cases of i integers; this means we only have to check for the points on the grid formed of integer coordinates. Nevertheless, having an analytical solution for $\hat{r}$ would be helpful in order to perform sensibility analysis on r^*, which could be done in future research.

A convenient feature of this graphical representation is that it allows us to analyse what happens if we change some parameters – for example, the environmental concern ω_j of outsiders or the technological levels σ_L and σ_H. Some examples of this are depicted in Figs. 2.12a–d, where some expected features can be observed.

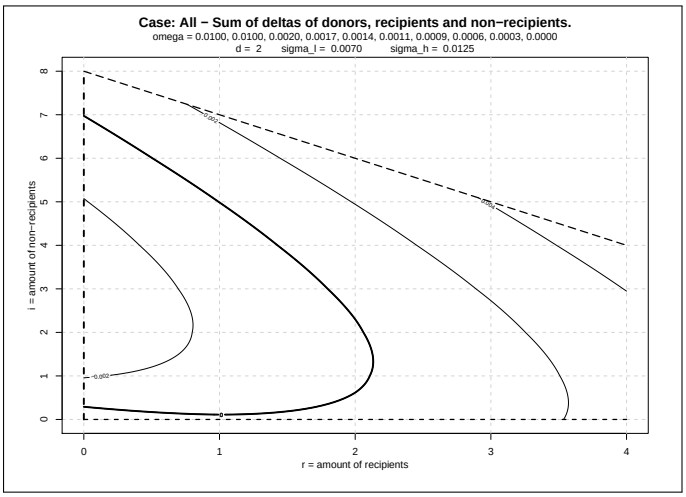
The lower the environmental concern of outsiders (lower values of ω_j), the harder it is to cross the mountain, meaning that the locus at which $\Delta\Pi^s_{oo} = 0$ switches to the right, and



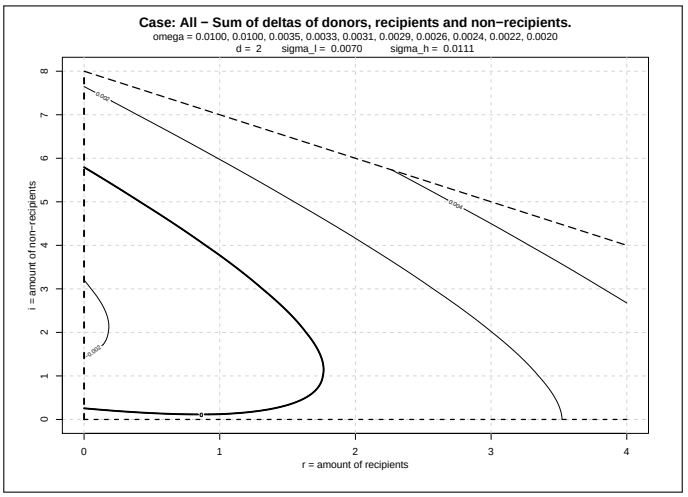
(a) Lower outsiders environmental concern.



(b) Same as (a), but even lower, with better good tech.



(c) Same as (b), but even lower environmental concern.



(d) More even σ_j with cheaper bad tech.

Figure 2.12: Some examples changing ω_j, σ_L and σ_H.

therefore $r*$ might increase. A similar effect occurs when making abatement technologies more expensive.

2.5 Conclusions

The present work explores a new approach for reaching an International Environmental Agreement (IEA) by combining two instruments, namely a technological transfer and a border tax. The idea is that the first instrument 'buys in' some countries to the initial coalition through technology transfers, following Heal and Kunreuther's (2011) idea that a set of countries could tip the rest to get to a clean equilibrium. The second one makes sure that these countries could impose costs such as border taxes on non-joiners in order

to persuade the rest of the countries to join the coalition, in a way similar to what happened with the Montreal Protocol and its subsequent amendments. Used alone, it might be the case that the border tax is not sufficient to induce the grand coalition, since it needs some critical mass to work. On the other hand, in some situations the transfer cannot work alone either, since it does not deal with the free-rider incentives. The results show that they may exhibit mixed effects when combined. They add to one another, but at the same time they erode their original persuading force. It turns out that depending on parameter conditions, the tax itself, the transfer itself, or both can induce the grand coalition.

I also analyse if non-signatories are willing to retaliate with a similar tax. I solve for which coalition sizes and parameter conditions non-members retaliate. I show that for small- and medium-sized coalitions, the retaliation is not worthwhile; for bigger coalitions it depends on specific conditions of possible deviators. In a recent paper by Nordhaus (2015), he explores, using a DICE model with 15 regions, the effectiveness of a border tax for inducing accession. However, he does not analyse the possibility of a retaliation tax. Hence, it could be explored, using a similar model, if non-signatories would be willing to retaliate. Taking all this into consideration, it could be the case that a big coalition, but not the grand coalition, becomes a stable solution. It is also worth noting that his model is applied to a specific set of parameters (the world divided into 15 regions) and the general result of the present work could help to study how sensitive his results are to changes in this configuration. On the other hand, it would be worth studying the feasibility of technology transfers, combined or not with a border tax, in a model like that.

Another interesting result is that in order to minimize transfers (if they are present in the optimal recipe) and improve the chances of reaching the grand coalition, the recipient countries are those with the least environmental marginal damage, also referred as to the least-green ones. Although it might be a coincidence, it looks like some relation exists with what happened with the bilateral agreement between the United States and China reached last November. Of course in this case, other considerations are at hand, such as the size and emissions levels of these two specific countries.

Finally, a further option can be suggested to expand this work. The strategic implications of being a recipient or not could be analysed. This comes from the fact that non-recipients do not receive any incentive (since they will accede due to the trade pressure). Therefore, this might induce them to join the agreement at an earlier stage and thus get the transfer. This strategic interplay can again be anticipated by the promoters of the IEA and by the rest of the players, enlarging the game set-up and eventually changing or re-asserting the previous result.

Bibliography

Lisa Anouliés. The strategic and effective dimensions of the border tax adjustment. *Journal of Public Economic Theory*, forthcoming. ISSN 1467-9779.

Scott Barrett. Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46:pp. 878–894, 1994. ISSN 00307653.

Scott Barrett. The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics*, 19(4):345–361, November 1997.

Scott Barrett. International cooperation for sale. *European Economic Review*, 45(10):1835–1850, December 2001.

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

Scott Barrett. Climate treaties and "breakthrough" technologies. *American Economic Review*, 96(2):22–25, 2006.

James A Brander, Barbara J Spencer, and Tariff Protection. Tariff protection and imperfect competition. In Henryk Kierzkowski, editor, *Monopolistic Competition and Product Differentiation and International Trade*, pages 194–206. Oxford Economic Press, 1984.

Carlo Carraro. The structure of international environmental agreements. In Carlo Carraro, editor, *International Environmental Agreements on Climate Change*, volume 13 of *Fondazione Eni Enrico Mattei (Feem) Series on Economics, Energy and Environment*, pages 9–25. Springer Netherlands, 1999. ISBN 978-90-481-5155-4.

Carlo Carraro and Domenico Siniscalco. Strategies for the international protection of the environment. *Journal of Public Economics*, 52(3):309–328, October 1993.

Claude D'Aspremont, Alexis Jacquemin, Jean Jaskold Gabszewicz, and John A. Weymark. On the stability of collusive price leadership. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 16(1):pp. 17–25, 1983. ISSN 00084085.

Effrosyni Diamantoudi and Eftichios S Sartzetakis. Stable international environmental agreements: An analytical approach. *Journal of public economic theory*, 8(2):247–263, 2006.

Thomas Eichner and Rüdiger Pethig. Self-enforcing environmental agreements and international trade. *Journal of Public Economics*, 102(0):37 – 50, 2013. ISSN 0047-2727.

Geoffrey Heal. Formation of international environmental agreements. In Carlo Carraro, editor, *Trade, Innovation, Environment*, volume 2 of *Fondazione Eni Enrico Mattei (FEEM) Series on Economics, Energy and Environment*, pages 301–322. Springer Netherlands, 1994. ISBN 978-94-010-4409-7.

Geoffrey Heal and Howard Kunreuther. Tipping climate negotiations. Technical report, National Bureau of Economic Research, 2011.

Michael Hoel and Kerstin Schneider. Incentives to participate in an international environmental agreement. *Environmental and Resource Economics*, 9(2):153–170, 1997. ISSN 0924-6460.

William Nordhaus. Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, 105(4):1339–70, 2015.

SJ Rubio and A Ulph. Self-enforcing agreements and international trade in greenhouse emission rights. *Oxford Economic Papers*, 58:233–263, 2006.

2.A Firm maximization problem

The firm's profit function to be maximized is the following:

$$\pi_j = \sum_{i=1}^{N} (1 - x^i - t_j^i - \sigma_j q_j) x_j^i \tag{A.1}$$

with $t_j^i = t$ if $i \in \mathcal{S}_k \wedge j \notin \mathcal{S}_k$, and $t_j^i = 0$ otherwise. I only consider situations where firms produce positive quantities in equilibrium. We can get the first order conditions, which are:

$$1 - x^i - t_j^i - \sigma_j q_j - x_j^i = 0 \qquad \forall i, j \tag{A.2}$$

Summing over i, conditions in equations A.2 can be rewritten as:

$$N(1 - \sigma_j q_j) - 2x_j - x_{-j} - t_j = 0 \qquad \forall j \tag{A.3}$$

where $t_j = \sum_{i=1}^{N} t_j^i$ is the sum of the tax rates 'paid' by a product produced by country j, $x_j = \sum_{i=1}^{N} x_j^i$ is the total output of firm j, and $x_{-j} = \sum_{k \neq j} x_k^i$ is the rest-of-the-world output. This can be re-written in a matrix form, getting:

$$\underbrace{\begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{bmatrix}}_{C} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{\overrightarrow{x_j}} = \underbrace{\begin{bmatrix} N - N\sigma_1 q_1 \\ N - N\sigma_2 q_2 \\ \vdots \\ N - N\sigma_N q_N \end{bmatrix} - \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}}_{D} \tag{A.4}$$

Using the Sherman-Morrison formula, we can get $\overrightarrow{x_j} = C^{-1} \cdot D$, being

$$C^{-1} = I - \frac{B}{N+1}$$

where I is the identity matrix and B (of dimension N by N) is a matrix of ones. This gives the general solution of:

$$\overrightarrow{x_j} = \frac{1}{N+1} \begin{bmatrix} N & -1 & \cdots & -1 \\ -1 & N & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & N \end{bmatrix} \begin{bmatrix} N - N\sigma_1 q_1 - t_1 \\ N - N\sigma_2 q_2 - t_2 \\ \vdots \\ N - N\sigma_N q_N - t_N \end{bmatrix}$$

Or equivalently:

$$x_j = \frac{1}{N+1} \left[N\Big(1 - N\sigma_j q_j - t_j + \sum_{i \neq j} \sigma_i q_i\Big) + \sum_{i \neq j} t_i \right] \tag{A.5}$$

Finally, for the country consumption levels x^i, we can use the FOC (eqn. A.2) and we get:

$$x^i = \frac{N - \sum_{j=1}^{N} \sigma_j \, q_j - t^i}{N+1} \tag{A.6}$$

with $t^i = \sum_{j=1}^{N} t^i_j$ being the sum of tax rates *imposed* by country i. Given that only signatories tax non-signatories' products, we have that:

$$t^i = \begin{cases} (N-k) \cdot t & \text{if } i \in \mathcal{S}_k \\ \\ 0 & \text{if } i \notin \mathcal{S}_k \end{cases}$$

which yields to

$$x^i = \begin{cases} \frac{N-\sigma_{\mathcal{S}}-(N-k)\cdot t}{N+1} & \text{if } i \in \mathcal{S}_k \\ \\ \frac{N-\sigma_{\mathcal{S}}}{N+1} & \text{if } i \notin \mathcal{S}_k \end{cases} \tag{A.7}$$

where $\sigma_{\mathcal{S}} = \sum_{j \in \mathcal{S}} \sigma_j$ (the sum of the coalition's marginal abatement costs). Replacing this result into the FOC (Eqn. (A.2)), we finally have the segmented production of firm j shipped to country i:

$$x^i_j = \begin{cases} \frac{1+\sigma_{\mathcal{S}}-(N+1)\sigma_j+(N-k)\cdot t}{N+1} & \frac{1+\sigma_{\mathcal{S}}-(k+1)\cdot t}{N+1} & \text{if } i \in \mathcal{S}_k \\ \\ \frac{1+\sigma_{\mathcal{S}}-(N+1)\sigma_j}{N+1} & \frac{1+\sigma_{\mathcal{S}}}{N+1} & \text{if } i \notin \mathcal{S}_k \end{cases} \tag{A.8}$$

$$\begin{array}{cc} \uparrow & \uparrow \\ \text{if } j \in \mathcal{S}_k & \text{if } j \notin \mathcal{S}_k \end{array}$$

Finally, having x_j, x^i and x^i_j and recalling the equations for the firm profit (π_j), consumer surplus (CS_j), environmental damage (DAM_j), and taxes collected (TAX_j), we get:

$$\pi_j = \begin{cases} \frac{N\cdot\aleph_j^2 + k(N-k)t\left(2\aleph_j+(N-k)t\right)}{(N+1)^2} & \text{if } j \in \mathcal{S}_k \\ \\ \frac{N\cdot\aleph_2^2 + k(k+1)t\left((k+1)t-2\aleph_2\right)}{(N+1)^2} & \text{if } j \notin \mathcal{S}_k \end{cases} \quad \text{with} \quad \begin{array}{l} \aleph_j = 1 + \sigma_{\mathcal{S}} - (N+1)\sigma_j \\ \\ \aleph_2 = 1 + \sigma_{\mathcal{S}} \end{array} \tag{A.9}$$

$$CS_j = \begin{cases} \frac{\left(N-\sigma_{\mathcal{S}}-(N-k)t\right)^2}{2(N+1)^2} & \text{if } j \in \mathcal{S}_k \\ \\ \frac{\left(N-\sigma_{\mathcal{S}}\right)^2}{2(N+1)^2} & \text{if } j \notin \mathcal{S}_k \end{cases} \tag{A.10}$$

$$DAM_j = -\omega_j \cdot \frac{(N-k)}{N+1} \cdot \Big((1+\sigma_{\mathcal{S}})N - k(k+1)t\Big) \tag{A.11}$$

$$TAX_j = t \cdot \frac{(N-k)}{N+1} \cdot \Big(1 + \sigma_{\mathcal{S}} - (k+1)t\Big) \tag{A.12}$$

2.B Proof of r^* being the least-green recipient group

As stated in section 2.4, the proof consists of showing that if conditions for *Cascading 1* hold, then conditions for *Cascading 2* hold too. In order to do so, I will use to following inequalities:

$$\underbrace{\underbrace{{}'\Pi_s^{d+r+i} \geq \overbrace{\Pi_s^{d+r+i}}^{\text{Cascading 1}} > \Pi_n^{d+r+i-1}}_{\text{Inequality 1}} \geq {}'\Pi_n^{d+r+i-1}}_{\text{Cascading 2}} \tag{B.1}$$

Hence, the proof shows that *Inequality 1* and *Inequality 2* hold. The first one says that the profit of an i^{th} country coming into the coalition ${}'\Pi_s^{d+r+i}$ (Cascading 2) is greater or equal than the same profit in the case of *Cascading 1*, where r' has been replaced by the original r and the i^{th} country entering the coalition might or might not be the same as *Cascading 2*, according to the following diagram:

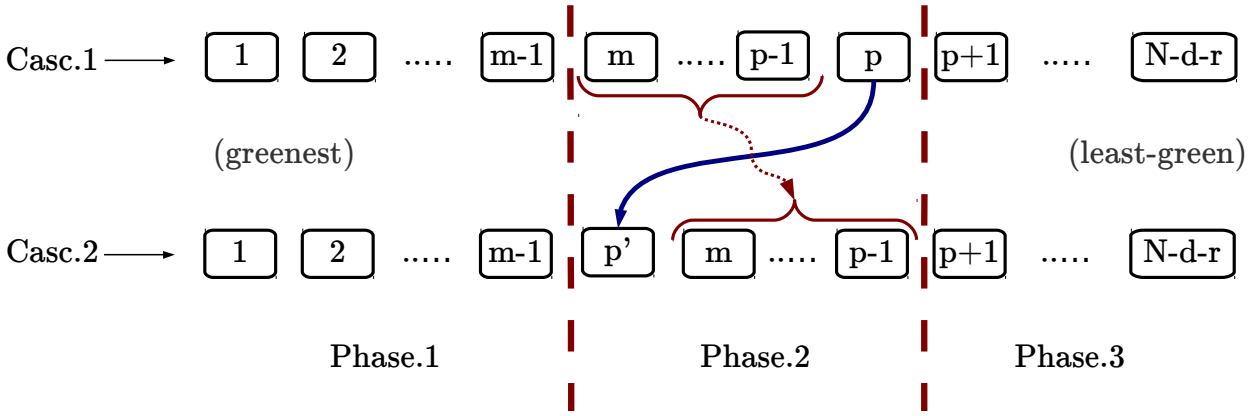


Figure 2.a: Non-recipients sequence for cascadings 1 and 2.

The same has to be done for *Inequality 2*, where now we have to show that the outside option of an i^{th} country coming into the coalition $\Pi_n^{d+r+i-1}$ (Cascading 1) is greater or equal than its counterpart in Cascading 2. To do so, I will analyse the possible changes in firm profits, consumer surplus, damages, and taxes collected. As shown in the previous picture, I will divide the analysis into three phases: 1, 2 and 3.

First, let us note that the amount $\sigma_{\mathcal{S}} = \sum_{j \in \mathcal{S}} \sigma_j$ (the sum of the coalition's marginal abatement costs) does not change between *Cascading* 1 and 2. This is due to the simple fact that the group of countries with good technology is always of size $(d + r)$. In the same way, $\aleph_j$ and $\aleph_2$ stay constant between those two cases. Therefore, the only difference that may arise comes from the replacement of ω_j, either of countries accessing the coalition

(the i's) or those in the recipient group (r or r').

Inspecting the firm profit equation (A.9), it is clear that its value does not change between *Cascading* 1 and 2. The same holds for the consumer surplus and for the taxes collected. Hence, we only have to focus on the damages coming from emissions. Studying this equation shows us that emissions are also invariant between these two cases, since they only depend on σ_S and k. Hence, the only changes comes directly from the term $-\omega_j$ (in Eqn. (A.11)).

Define $\Delta D_1 = {}'\text{DAM}_j^{d+r+i} - \text{DAM}_j^{d+r+i}$, which is the difference in damages of country j between *Cascading* 2 and 1 cases, in the presence of a coalition of size $(d+r+i)$. In the same manner, define $\Delta D_2 = {}'\text{DAM}_j^{d+r+i-1} - \text{DAM}_j^{d+r+i-1}$, which is just the same, with a coalition of size $(d+r+i-1)$. Finally, denote ω_p the corresponding marginal damage for the country p in r being replaced by a less-green country q in r'. And let ω_q be the marginal damage of the replacing country q.

Let us start with the signatories (Inequality 1). In phase 1, $\Delta D_1 = (\omega_p - \omega_q) \cdot \textit{emissions}$, which is positive. In phase 2 we will have a similar case, where the pair of ω's at stake will be: ω_p v/s ω_m, ω_m v/s ω_{m+1} ... ω_{m+j-1} v/s ω_{m+j} (recall Fig. 2.a) and therefore resulting again in $\Delta D_1 > 0$. For phase 3, since the coming countries i's and the group already in the coalition $(d+r+i-1)$ have the same ω's, $\Delta D_1 = 0$. Therefore, for phases 1, 2 and 3 together we have that $\Delta D_1 \geq 0$, which proves Inequality 1.

For the case of non-signatories (Inequality 2) we have that for phases 1 and 3, $\Delta D_2 = 0$, since here we are only concerned on the entering country. For these two phases, the entering country is the same for both *Cascadings*. Following the same reasoning of phase 2 in the previous paragraph, we get that $\Delta D_2 < 0$ for this phase. Putting the three phases together leads to $\Delta D_2 \leq 0$, which again proves Inequality 2 (note that Inequality 2 has the *Cascading* 1 and 2 inverted with respect to Inequality 1).

This last point proves that *Cascading* 1 implies *Cascading* 2. We iterate on the swapping process in r' until the recipient group becomes r^*, the least-green outsider, which finishes the proof.

Chapter 3

The Evolution of a "Kantian Trait": Inferring from the Dictator Game

3.1 Introduction

The goal of this paper is twofold. First, using the transmission theory of social traits, specifically the population dynamics tool-kit, I develop the transmission dynamics of a trait that lies in a continuous segment. This trait is transmitted using some myopic preference of parents (concerning their offspring utility) as in Bisin and Verdier (2000). Within this set-up, the literature usually develops to find the equilibrium. I invert this question and ask: Which parents' preference function would lead to a given population distribution? I solve the conditions that have to be met and I also develop an algorithm to solve for more complex cases. On the other hand, and using the previous results, I connect the transmission theory of social traits with the results of a well known experiment in economics, the Dictator Game (DG). In order to do so, I use the experimental results of the DG in order to infer the distribution of a moral trait in a society. In this case, the moral trait is what I call a 'Kantian morale'. It is possible to map the responses of DG experiments into these types of moral traits. Assuming that distributions of actual societies are the result of a long evolutionary transmission process of these kind of traits, and using the aforementioned results, I ask if there is such an evolutionary process that could explain the results of DG experiments. It turns out that such a process could exist (mathematically speaking), and if this is the case, it would imply that homo-oeconomicus parents have stronger feelings about having offspring similar to them, than their Kantian counterparts. This result can turn out to be important when dealing with environmental challenges, where individual provision of environmental public goods and coordination are greatly needed in order to sustain a clean environment.

The literature on social trait transmission uses the population dynamics tool kit in

order to model how the next generation will be, according to the present state of the distribution of traits and the 'forces' involved in the evolving process. For example, Bisin and Verdier (2000) suppose that agents of the present generation, using a myopic empathy when considering their offspring's utility, try to transmit their own traits to their children.[1] This myopic empathy means that parents evaluate their offspring's utility according to their own. This implies that parents will evaluate their children's actions using their own utility functions, which in turn will yield to a lower utility of their offspring, from the parents' point of view, if their traits differ from those of their parents. I use this starting point and I make two modifications: I assume there exists a 'myopic dislike' function $v(\cdot)$ of the agent towards his or her child's trait when this trait is different from their own; and I assume that this trait can be modelled with one variable positioned in a continuous line between zero and one. This departs from the literature, where usually authors study a specific scenario with a finite number of types of agents, usually being equal to two or three.

On the other hand, the Dictator Game (DG) is one of the simplest and most replicated economic experiments in the game theory area. In a nutshell, the experiment consists of recruiting two people and giving an amount of money (or another valuable thing) to one of them, chosen randomly. The person who received the money is called the Proposer (or Dictator) and he is asked to share some fraction (or none) of this money with the second person, called the Responder. After this, the money is split according to this decision. Therefore, the Responder has no say in this game; we are only interested in the Proposer's decision. According to the homo-oeconomicus theory, the Dictator should never give a dime, but these experiments show a constant and considerable amount of people sharing some part, even to a 50/50 proportion or more. Different explanations of this phenomenon exist. Assuming that the Proposer has no direct or indirect relationship with the Responder, which is usually the experiment set-up, these explanations point to the idea of a social norm and/or a moral trait, both of which are transmitted between generations. There are other experiments (games) that could be used in order to infer traits like this one. I have chosen the DG since it gives answers pointing toward a single motivation (or the closest we can get to this idea). Other games that we could test are the Ultimatum Game (UG) and the Public Good Game (PGG). In the former, which is a DG where the Responder can reject the Proposer's offer, the latter will try to choose an offer that the Responder will accept. This is usually related to social norms and therefore, even for the full homo-oeconomicus, it would be never optimal to choose zero give rate. In other words, the results of this game provide information about both how Kantian (or a similar trait) the person is and the **expected** norms that the Proposer foresees. Since

[1] This type of trait transmission has also been used in Bisin and Verdier (2001), Hauk and Saez-Marti (2002) and Sáez-Martı and Zenou (2012).

there is no clear way to separate this from the data (or at least, no easy way) I preferred to discard the use of the UG. In the PGG, players are endowed with tokens that they can choose to contribute to a pool. The sum of all tokens collected is multiplied by a factor greater than one and less than the number of players, and this "public good" payoff is evenly divided among players. Each player also keeps the tokens they do not contribute. In this case, it is possible to extract similar information, at least in the first shot of the game. If the game is repeated (especially using the same group of people, which is the usual set-up), then we will observe other things such as, for example, how cooperation can develop (like in a tit-for-tat strategy). Hence we run into a similar issue as we do with the UG. For these reasons I use the DG, although other approaches could be used as extensions to study different traits.

Following the idea developed in Chapter 1 regarding a 'Kantian morale', where agents are endowed with a Kantian trait that lies in a continuous spectrum, I can map the DG responses to a Kantian trait level. In this framework, I define a (fully) Kantian person as an agent maximizing his utility *assuming* that everyone acts as he does, in contrast with the homo-oeconomicus agent, who just maximizes his utility in a selfish manner. It is important to note that this is **not** what Kant meant in his seminal work. I use the term 'Kantian' in the same way as Laffont (1975). The idea here is that each person wants to behave in a morally responsible way, at least to some degree. Therefore, if the degree is full, I call this agent a fully Kantian person. If the degree is null, we have the fully homo-oeconomicus case. In this interpretation, as in Laffont's, the fully Kantian person is making his choice under the assumption that everyone behaves as he does. Of course, he or she does not expect for this to happen; it is just the empirical way of finding the 'good thing to do'. Of course, this is not what the categorical imperative is about, but I borrow the name since is a close practical implementation of Kant's idea. Degrees in between reflect people who give some weight to the Kantian responsibility (which I call α) and the rest to the homo-oeconomicus way of behaving (consequently weighted by $(1-\alpha)$). Therefore, we can have people behaving in an intermediate manner, which is captured by this continuous Kantian trait. While evidently this is not exactly what Kant meant in his seminal work, it attempts to reflect what we observe in real life; for example, using DG experiment results gives us a great variety of outcomes. An interesting result of the DG experiments is that there is a kind of polarization of the distribution, having roughly one third of the population acting in a purely homo-oeconomicus way (giving nothing), another third or so sharing half of the pie (or even more), which I translate to being fully Kantian, and the rest of the people distributed, almost uniformly, between these two extremes.

With these two ingredients in mind, the objective of this paper is to rationalize the

evolution of the distribution of the Kantian trait to one that could account for the results of DG experiments.[2] In order to do so, I will first develop a continuous model of population dynamics, using the discrete one as a starting point. I assume that there exists a $v(\cdot)$ function that represents the parent's 'dislike' or loss in utility of having a child with a trait different from their own. The input of the function is the difference of the child's trait compared to his or her parent's trait. A positive amount means that the child is more Kantian than his or her parent, where a negative one means the inverse case. The $v(\cdot)$ function does not need to be symmetrical in zero. Actually, we will see that in order to match the DG experiment's results, it will not be symmetrical. With this, I find equilibrium conditions of the dynamics where the distribution of the population stops evolving. I am assuming here that the results of the DG experiments are actually in equilibrium, or quite close to it. With this set-up, I show some simple results concerning possible solutions of $v(\cdot)$ functions, which evolve into a given equilibrium distribution of the population. This distribution lies in the continuous segment $[0, 1]$, and it can have points of high concentration, modelled with Dirac deltas. The only condition for the distribution is that its integral has to be equal to one, as in any distribution.

Having the basic results of the equilibrium conditions and some properties of $v(\cdot)$, I move into finding $v(\cdot)$ function(s) that can make a population distribution evolve into what we have as results in DG experiments. It turns out that solving this problem analytically is not plausible for this asymmetric case, and I have to rely on simulations. Here I develop an algorithm that takes the resulting (final) distribution as input and an initial guess for $v(\cdot)$; using an iterative process, it converges to a $v(\cdot)$ function that meets the equilibrium conditions. The chapter starts in Section 3.2 with the basic discrete model and explains how I transform it to its continuous counterpart. This section also includes some basic properties of the solutions. In Section 3.3, I move to the general case, introducing the algorithm and showing the solution for the case of the DG experiments' results. Section 3.4 concludes.

3.2 The Initial Model

3.2.1 The starting point

I begin with the trait transmission, using as a starting point the model used by Bisin and Verdier (2000). They used what it is called 'vertical' and 'oblique' cultural transmission. The idea is that parents will try to transmit their own trait to their children, which

[2]The word 'evolution' used here has its origins in biological studies, where biologists analysed how different animal populations 'evolve' depending on the presence of other species. It does not refer to genetic evolution, but since the processes are close in essence, it inherited the name.

will be effectively transmitted with some probability τ. If they fail to transmit the trait, then with probability $(1-\tau)$ the child is matched randomly with an individual of an old generation and adopts his trait. In their paper, the authors deal with a population consisting of only two types of people, therefore having the vertical transmission probabilities of τ^a and τ^b (one for each type). By calling q_t the share of population of type a in period t (and therefore having $(1-q_t)$ the share of population of type b), the transition probabilities P_t^{ij} of type i having a child of type j, are easily calculated (as in Bisin and Verdier (2000)):

$$P_t^{aa} = \tau^a + (1-\tau^a)q_t \qquad P_t^{ab} = (1-\tau^a)(1-q_t) \tag{3.1}$$

$$P_t^{bb} = (1-\tau^b)(1-q_t) \qquad P_t^{ba} = (1-\tau^b)q_t \tag{3.2}$$

Given these probabilities, the share of type a in period $t+1$ is derived too:

$$q_{t+1} = q_t + q_t(1-q_t)[\tau^a - \tau^b] \tag{3.3}$$

This is the standard evolution equation found in population dynamics, as in for example Sigmund (1986), Silverberg (1997), Hofbauer and Sigmund (2003), and Harper (2009). Still following Bisin & Verdier's (2000) paper, the parent will bear a cost when socializing their child with a given trait. In this case, this cost is denoted by $H(\tau^i)$, depending on the socialization effort τ^i. The parent chooses τ^i that maximize

$$\beta[P_t^{ii}V^{ii} + P_t^{ij}V^{ij}] - H(\tau^i) \tag{3.4}$$

where β is the discount rate and V^{ij} is the utility of a child of type j perceived by a parent of type i. Again following the literature, I assume that parents act according to 'imperfect empathy', meaning that they evaluate their child's utility through their own imperfect lenses. This is the point where I depart from the literature. First, I assume that V^{ij} is constant in time and well-known by the parents of type i. These values do not depend on the composition of the society, as they do in some cases in the literature. In any case, the maximization in Eqn. 3.4 does depend on the society composition, through the transition probabilities P_t^{ij}. Secondly, I normalize the value of V^{ij} such that instead of being the child j's utility viewed in parent i's eyes, it will be the *difference* of this aforementioned utility and the parent's. In other words, I define $V(i,j) = V^{ij} - V^{ii}$. Therefore $V(i,j)$ is the loss of utility that a parent of type i has, when having a child of type j. This will turn out to be handy in the generalization to n types of agents.[3]

[3]This modification does not change the maximization problem. It is easy to verify this by replacing the term P_t^{ii} with $1 - P_t^{ij}$ in Eqn. (3.4). Since V^{ii} is constant with respect to τ^i, we have the same maximization program. In the general case, we make $P_t^{ii} = 1 - \sum_{j \neq i} P_t^{ij}$.

Returning to the population dynamics and observing Eqn. (3.3), it is easy to see that if both types coexist ($q \neq 0 \wedge q \neq 1$), then the system stops evolving when $\tau^a = \tau^b$, meaning that both types of parent are exerting the same amount of effort in socializing their children.

3.2.2 Extending the model

Now I will add more types to the model, then transform it into one with types of people lying in a continuous segment. First, we have n types of agents; let us have them ordered, as for example with the natural numbers. In other words, we have different degrees of a trait and the n types just signify the strength of this trait. As explained in the Introduction, I will relate this type i with how Kantian or homo-oeconomicus a person is, as I did in Chapter 1. Hence, we can think of having n types $1, 2, \ldots, n$, where $i = 1$ means a fully homo-oeconomicus person and $i = n$ means a fully Kantian one. Those in between are ordered in the sense that if $j > i$, it means that type j is more Kantian than i, although this idea can be applied to any trait that allows an ordering.

I also assume that $V(i, j)$ is a 'dislike' function, as previously mentioned, and that it increases with the difference of i and j. This means that the more different a child turns out to be (with respect to his parent), the bigger the disutility that his parent bears. We also have, following its definition, that $V(i, i) = 0$, $\forall i$.

Following the previous notation, let q_t^i be the share of type i in the population at time t. It is now easy to extend the two type equations (Eqns. (3.1) and (3.2)) into n types. We can now write the transition probabilities:

$$P^{ii} = \tau_i + (1 - \tau_i) q_t^i \qquad P^{ij} = (1 - \tau_i) q_t^j \qquad \forall i \neq j \tag{3.5}$$

We can also construct a matrix V with its elements being: $V^{ij} = V(i, j)$ and the effort vector $\vec{\tau} = (\tau_1, \tau_2, \ldots, \tau_n)^T$. The maximization problem for each agent j will be:

$$\max_{\tau_j} \beta \Big(\sum_{i=1}^{n} P^{ji} \cdot V^{ji} \Big) - H(\tau_j) \tag{3.6}$$

Noting that the diagonal of V is full of zeros (since $V^{ii} = 0$), we can redefine the matrix P without changing the system of equations by having $P = (\vec{1} - \vec{\tau}) \cdot \vec{q}^T$, where $\vec{1}$ is a vector of ones and $\vec{q}$ is the vector composed by the shares of each type (and hence, the sum of its

elements equals one). With this, the solution of the maximization problem is:

$$\beta \frac{\partial((P \cdot V^T)_j)}{\partial \tau_j} = -\beta \sum_{i \neq j} q_t^i \cdot V(j,i) = H'(\tau_j) \tag{3.7}$$

For simplicity, let me assume that $H(\tau) = 1/2\tau^2$, and since $V(j,i)$ is non-positive, that $V(j,i) = -c \cdot v(j,i)$, with $c > 0$ constant and $v(j,i) \geq 0$. There are two things to say about the first assumption. First, I suppose that the cost function $H(\cdot)$ is the same for everyone. Now, since the idea is to find the force behind the trait, I focus on $V(\cdot)$ rather than $H(\cdot)$. While it could also be the case that the cost is related to the trait of the parent, I think assuming these two are not related (or are only loosely related) is a safe course. Furthermore, I want to investigate the willingness of a parent to have a child with a trait close to their own ($v(j,i)$ function) more than the cost function of a parent's type. Second, I assume a specific functional form for $H(\cdot)$, which is just to make computations easier. In fact, I only need $H(\cdot)$ to be strictly increasing in order to have the same equilibrium condition. This will become clear in the following paragraphs. With all these assumptions, we can write the solution of the maximization problem:

$$\tau_j = \beta \cdot c \sum_{i \neq j} q_t^i \cdot v(j,i) \tag{3.8}$$

We can now return to the dynamic of the population. It is easy to show that the general solution the evolution of q_t^j is:[4]

$$\Delta q_t^j = q_{t+1}^j - q_t^j = (\tau_j - \overline{\tau}) \cdot q_t^j \qquad \text{with} \qquad \overline{\tau} = \sum_{i=1}^{n} \tau^i q_t^i \tag{3.9}$$

It is worth noting that any constant factor that shows up in Eqn. (3.8) will also appear in $\overline{\tau}$ and therefore in Δq_t^j. This will only change the speed of the evolution process; the equilibrium will be the same. Following this, we do not need to worry about the constant factors β and c in Eqn. (3.8). From Eqn. (3.9) it is straightforward that at equilibrium, when $\Delta q_t^j = 0$, for all surviving types j (i.e. $q_j^* > 0$), $\tau_j = \overline{\tau}\ \forall j$. In words this means that the transmission effort τ_j that, for those types that did not disappear in the evolution process ($q_j^* > 0$), is the same (or constant). This is the equilibrium condition.

The continuous case

The idea is to extend the model to one with continuous types of agents. Now agents' type α will lie in the continuous segment $[0,1]$. Therefore, the distribution of the popula-

[4] For details, please see Appendix 3.A.

tion will be $f(\alpha) \geq 0$, with

$$\int_0^1 f(\alpha) d\alpha = 1 \tag{3.10}$$

The condition that $\tau_k = \overline{\tau}$ for all types of agents can be rewritten as following, using the result of Eqn. (3.8):

$$\int_0^1 v(j,\alpha) f(\alpha) d\alpha = \tau(j) = \text{constant} \qquad \forall j \mid f(j) > 0 \tag{3.11}$$

In other words, this means that a population whose dynamic is 'defined' by the function $v(\cdot)$ will converge into a distribution $f(\alpha)$ when condition (3.11) is met. As with other results in population dynamics, there may exist more than one converging distribution $f(\alpha)$ for a given dynamic, and the final distribution of the population depends on the initial state of the population distribution (see for example Zeeman (1980) and Friedman (1998)). Another way to use condition (3.11) is to ask: Which $v(\cdot)$ dynamics (if there are any) would converge to a population defined by $f(\alpha)$?

In order to start answering this question, let us try two simple examples. First, let us answer this question with the following $f(\alpha)$:

$$f(\alpha) = \frac{1}{2}\,\delta(0) + \frac{1}{2}\,\delta(1) \tag{3.12}$$

where $\delta(x)$ is Dirac delta function[(5)] centred in x. This means that $f(\alpha)$ is a polarized population in which half of the people are concentrated in $\alpha = 0$ and the other half in $\alpha = 1$. In order to simplify the formulation, note that $v(j,\alpha)$ is actually a function of the difference of j and α. Hence, we can rename it $v(j-\alpha)$, which will be handy later on. It turns out that condition (3.11) simply becomes $v(j) + v(1-j) = constant$, for $j = 0$ and $j = 1$. This translates to $v(0) + v(1) = v(1) + v(0)$, which is always true, no matter which function $v(\cdot)$ we use. This means that this polarized distribution is a solution of **any** $v(\cdot)$ dynamics.

Let us now try with a more complex example:

$$f(\alpha) = C_1\,\delta(0) + C_2 + C_1\,\delta(1) \qquad \text{with } 2C_1 + C_2 = 1 \tag{3.13}$$

In this one we have also a (semi)polarized distribution, where there are people in the intermediate values of α. These intermediate people are distributed uniformly. Solving

[(5)]Dirac delta function or δ function is zero everywhere except at zero, with an integral of one over the entire real line.

again for condition (3.11)[6], we get the following:

$$C_1(v(j)+v(1-j))+C_2\Big(\int_0^j v(j-x)dx+\int_j^1 v(x-j)dx\Big)=\text{constant} \tag{3.14}$$

Functional solutions for this integral equation can be hard to find. A way to find (at least) some of them would be to transform it into a differential equation and try to find some solutions in that domain. By calling $w(j)=\int_0^j v(x)dx$, and hence $w'(j)=v(j)$, we can transform the previous equation into:

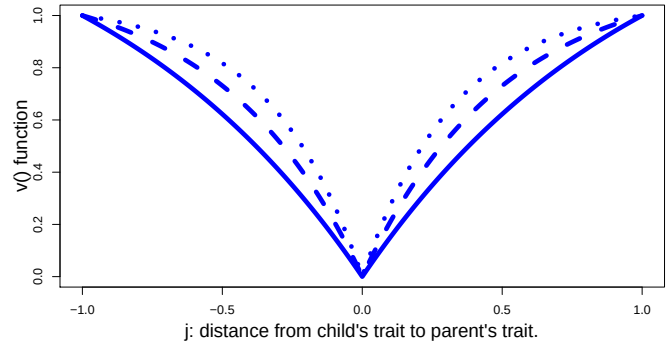
$$C_1\cdot\big(w'(j)+w'(1-j)\big)+C_2\cdot\big(w(j)+w(1-j)\big)=\text{constant} \tag{3.15}$$

Since this particular distribution $f(\alpha)$ is symmetric, I focused on searching for solutions that are symmetric around zero, meaning that $v(-j)=v(j)$. Two solutions for this equation are:[7]

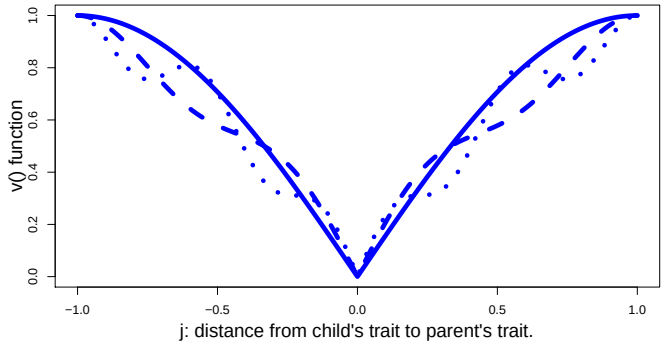
$$v(j)=K-a\cdot c\cdot e^{-a|j|} \qquad \text{with} \quad a=C_2/C_1 \ \wedge \ K=a\cdot c \tag{3.16}$$

$$v(j)=b\cdot sin\Big(\frac{(4k-3)\pi}{2}\cdot|j|\Big) \qquad \text{with} \quad b>0 \text{ (see footnote (7))} \ \wedge \ k\in\mathbb{N} \tag{3.17}$$

Constant K in (3.16) is such that $v(0)=0$, and c is an arbitrary positive number. Taking these solutions into account, we now have a family of solutions, just by making a positive linear composition of different cases of (3.16) and (3.17). It is worth noting that in the case of the sine solution (3.17), when adding a higher-order sine (higher values of k), one should pay attention to its factor, since we want the final $v(j)$ to be an increasing function. In Fig. 3.1, some examples for both cases are plotted.



(a) Examples of solution in (3.16).



(b) Examples of solution in (3.17).

Figure 3.1: Some examples of solutions for $v(j)$.

[6] For details, please see Appendix 3.B.

[7] The parameter b in solution (3.17) is linked to C_1, C_2 and k, and therefore, this is a solution for a specific combination of C_1 and C_2. In order to have a generic solution for any C_1 and C_2, a linear combination of sines is needed (with different values of k). For details on this point and on the development of the solutions, please refer to Appendix 3.B.

3.3 General Solution

It is easy to see that finding an algebraic solution for any given distribution $f(\alpha)$ is not possible, and therefore I will rely on simulations and a numerical solution. In order to do so, I propose an algorithm that finds a solution, for a given distribution $f(\alpha)$. The algorithm starts from an initial guess for $v_0(j)$ and converges to a solution of $v(j)$. As expected, the solution found will depend on the initial guess.

The algorithm is the following:[8] take an initial guess of $v(j) = v_0(j)$. Compute $\tau_0(j) = \int_0^1 v_0(j-\alpha)f(\alpha)d\alpha$ for $j \in [0,1]$ (which is the expression in (3.11)). If $\tau_0(j)$ is constant[9] (for $0 \leq j \leq 1$), then $v_0(j)$ is a solution. If not, define an *adjustment function* as:

$$a_0(j) = \frac{1}{\tau_0(1-2j)} \tag{3.18}$$

and compute a new guess function $v_1(j) = a_0(j) \cdot v_0(j)$. Compute the new $\tau_1(j)$ and restart the process.

The intuition for choosing this adjustment function is the following: We want to find a $v(j)$ function that, when plugged into the equilibrium condition (Eqn. (3.11)), gives us a constant value for $\tau(j)$. Now, when moving j between zero and one, this integral (the equilibrium condition) is the multiplication of $f(\alpha)$, and $v(j-\alpha)$ that 'moves' along it. Therefore, if we do not get a constant value for $\tau(j)$, we want to correct $v(j)$ such that it does. The multiplicative inverse accounts for this purpose. As to for the factor 2 in j, this is due to the fact that the range of $v(j)$ is between -1 and 1, where it is only $[0,1]$ for $\tau(j)$. Finally, the $(1-2j)$ term, which is just a vertical mirror of the function (expanded by 2, as just explained), comes from the fact that we are making this 'sweep' in the inverse direction (note α in $v(j-\alpha)$, inside the integral in condition (3.11)).

It is worth noting that it would be desirable to search $v(j)$ solutions with a particular functional forms. Unfortunately this can strongly depend on the input function $f(\alpha)$ and therefore, using this new constraint, it might be the case that the algorithm does not converge (probably meaning that this particular functional form cannot be a solution for that particular input $f(\alpha)$). Due to this issue, I leave the algorithm to search for a general solution. One drawback to this method could be that it might find multiple solutions, as

[8] Properties of the algorithm can be found in Appendix 3.C. I show that if at some iteration k, $v_k(j)$ is a solution, the algorithm converges, and that if it converges, then $v_k(j)$ is a solution. I do not show that the algorithm will always converge, although simulations' results suggest it does. Of course, if there is no solution, the algorithm will not converge. The algorithm was programmed in R language and is available upon request.

[9] Or close enough to be constant, depending on the desired precision.

we will see in the following pages. Nevertheless, it could be interesting to explore this first idea when solving for particular cases of $f(\alpha)$.

3.3.1 A simple application

With this algorithm it is possible to find solutions for more complex distribution functions – in particular, cases where the distribution is not symmetric. Returning to the initial goal, we can find $v(j)$ functions that account for the evolution of a Kantian trait that can explain, at least in part, the behaviour found in the Dictator Game experiments. Different papers can be used as sources of information, and I focus on Engel (2011), which is a meta study on Dictator Games. It is useful for my purpose since it compiles a large amount of experiments and responses. In Fig. 2 of this work, he shows the result of 328 treatments with full range information, composed of the answers of 20,813 people. I replicate this in Fig. 3.2a in the coming pages.

Recalling the Introduction, the DG is played between two agents, who will be chosen randomly to be the Proposer (Dictator) and Responder. We are interested in the Proposer's action,[(10)] and each agent has a 50% chance to be elected as such. He or she will choose to give a share of the money they receive when chosen Proposer. Therefore, in order to use this information, I will transform the DG responses (give rate) into what I call Kantian trait. As stated in the Introduction and following the discussion in Chapter 1, I can associate a Kantian 'measurement' to each respondent based on his give rate. Therefore, each person is defined by a α value (his Kantian trait) when he is acting in a way that maximizes the following utility function:

$$U(\cdot) = (1-\alpha)\cdot U_H(\cdot) + \alpha \cdot U_K(\cdot) \tag{3.19}$$

where $U_H(\cdot)$ is the utility function for an homo-oeconomicus agent and $U_K(\cdot)$ is the one for the Kantian person. A Kantian person is defined as one who maximizes his utility *assuming* that everyone acts as he does. One useful feature of this approach is that it transforms different experiments' results into a single measurement of what we could call a Kantian trait. In the case of the DG, the utility function is a function that transforms money (or the asset used in the DG) into utility. Typically this is a consumption utility function, with the classic properties of being increasing and concave. Using different utility functions, one can map different give rates (between zero and one half) into Kantian traits α, which will lie in the segment $[0, 1]$. To see this, let $u(\cdot)$ be the consumption utility function in the DG experiment, γ the fraction shared by the Proposer (give rate), and C the amount received by this agent. Therefore we have, for the homo-oeconomicus agent,

[(10)] The Responder has no say in this game.

the fully Kantian agent, and the general case, the following maximization programs:

$$\max_{0\leq\gamma\leq 1} \underbrace{\frac{1}{2}u((1-\gamma)C)}_{U_H(\cdot)} \quad \rightarrow \quad \gamma^* = 0 \tag{3.20}$$

$$\max_{0\leq\gamma\leq 1} \underbrace{\frac{1}{2}u((1-\gamma)C) + \frac{1}{2}u(\gamma C)}_{U_K(\cdot)} \quad \rightarrow \quad \gamma^* = 1/2 \tag{3.21}$$

$$\max_{0\leq\gamma\leq 1} (1-\alpha)\cdot\frac{1}{2}u((1-\gamma)C) + \alpha\cdot\left[\frac{1}{2}u((1-\gamma)C) + \frac{1}{2}u(\gamma C)\right] \tag{3.22}$$

The $1/2$ value comes from the chance of being chosen as Proposer or Responder. Therefore, for the homo-oeconomicus case, he or she only evaluates when they are chosen as Proposer. From the point of view of the Kantian person, who decides assuming that everyone will act as they do there are gains when being a Proposer (with one half chance) and Responder (the other half), hence the formulation in (3.21). For the first two cases we have two straight solutions, no matter which utility function $u(\cdot)$ we use. For the homo-oeconomicus agent ($\alpha = 0$), the give rate is zero ($\gamma^* = 0$), as in Eqn. (3.20). For the fully Kantian agent ($\alpha = 1$, Eqn. (3.21)), the give rate is equal to one half ($\gamma^* = 1/2$). As for those agents with $0 < \alpha < 1$, the value of γ^* that solves the maximization program, in Eqn. (3.22), is:

$$u'(\gamma C) = \alpha u'((1-\gamma)C) \tag{3.23}$$

Therefore, the relationship between α and the pair (γ, C) will depend on the choice of $u(\cdot)$, as in the following examples:

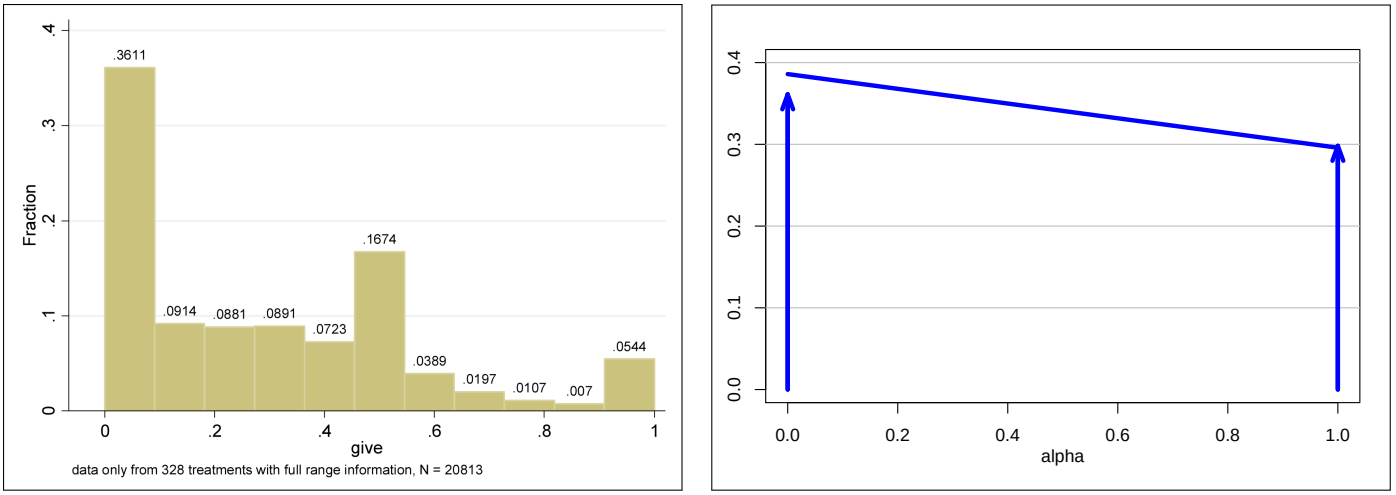
$$u(c) = ln(c) \quad \rightarrow \quad \alpha = \frac{\gamma}{1-\gamma} \tag{3.24}$$

$$u(c) = \frac{1}{1-\epsilon}c^{1-\epsilon} \quad \rightarrow \quad \alpha = \left(\frac{\gamma}{1-\gamma}\right)^{\epsilon} \tag{3.25}$$

In these cases, I used a Constant Relative Risk Aversion (CRRA) utility function, with ϵ as the measure of risk aversion ($\epsilon = 1$ in the first case). Given this special form of the utility function, we find that the relationship does not depend on the amount of money to share, but only on give rate. In general, give rates do not substantially change with the amount to share, except when it comes to big ranges in the amount to divide, as in Novakova and Flegr (2013).[11]

[11] Usually DGs are played with different amounts of money involved, although there is not normally a substantial difference among these values. The authors investigate a bigger difference, ranging approximately between \$1 and \$10,000, although their study is a survey and no real money was actually provided.

Concerning the value of ϵ, figures too different from $\epsilon = 1$ will produce a loss of information (when transforming from γ to α). With this consideration, and bearing in mind that the log function has been widely adopted in the literature, I use an ϵ equal to one.[12] Finally, for those people that gave more than one half, we should ask ourselves if we should discard them when transforming the distribution to a Kantian equivalent or, as I do, assume that these people are fully Kantian and the extra give comes from either miscalculation, misunderstanding, another motivation, or a combination of these factors. This assumption is in line with some detailed results found in O'Garra and Krantz (2014),[13] as in the follow-up questions (which were open-ended) for dictators. These questions were aimed at identifying the reasons for their choices and are in line with my previous statement. In any case, if we were to discard this information, the main result does not change much (results are included in Appendix 3.D). In Fig. 3.2 below the original results of Engel and its transformation to Kantian measurement are plotted.[14] The



(a) Distribution of give rates, from Engel (2011).

(b) Distribution in Kantian measurement.

Figure 3.2: Results from DG meta study and its equivalent in Kantian measurement.

two arrows in $\alpha = 0$ and $\alpha = 1$ in Fig. 3.2b are Dirac delta functions representing the two groups of people that cluster in these values, where the straight line is the density function of people with a Kantian measurement α in between. To arrive at this distribution, we have to find the parameters of this straight line (intercept and slope) that, when converted to their equivalent give rates γ, best fit the data. It turns out that a quasi-linear decreasing distribution fits the data quite well (which was suggested by the original histogram in

[12] For some examples and a deeper explanation, please see Appendix 3.E.

[13] I greatly thank Tanya O'Garra of the Center for Research on Environmental Decisions, Earth Institute, Columbia University for providing me with the data from her Dictator Game study.

[14] I greatly thank Professor Dr. Christoph Engel of the Max Plank Institute for Research on Collective Goods, for providing me with the data of his meta study on Dictator Games.

Fig. 3.2a); this distribution does not vary much when we use different transformations between α and γ. The slope and intercept will change a bit, maintaining a decreasing trend. On the other hand, the cluster of people giving zero are translated into the Dirac delta in $\alpha = 0$, and people giving 50% or more are transformed into the Dirac delta in $\alpha = 1$. These two magnitudes are, of course, invariable with respect to the previous transformation. In the case of the depicted distribution of α (Fig. 3.2b), I used a simple linear relationship, although different transformations lead to very similar results.

Before discussing the solution, it is worth recalling two assumptions used to apply the previous results and algorithm. First, I have assumed that the implicit utility function $u(\cdot)$ is increasing and concave. Concavity is needed in order to have a fixed point for the fully Kantian agent (result in Eqn. (3.21)), which in this case is a give rate of $\gamma^* = 1/2$. Since this is a fairly standard assumption, I do not see the need to comment further on it. The second assumption has to do with the equilibrium condition of the society. I have supposed that the society is at its long-term equilibrium. This is certainly not necessarily the case, but I assume it is for two reasons. The first has to do with simplicity. If I do not assume that the society is at equilibrium, then I should consider not only the distribution of the trait, but also the speed of change. This term is in turn affected by the factors β and c (Eqn. (3.8)). At this point it becomes clear that to determine these parameters, we would need to perform more experiments (most likely with the same people), but the initial goal was to use data readily available from previous experiments and publications. The second basis has to do with the idea that, at least concerning a Kantian trait, if societies are not at equilibrium they are probably close to it. This is (most likely) more true in undeveloped countries that have little contact with the rest of the world than in places with a rotating population. Still, if the society is close to equilibrium, one can see that the previous idea should also work.

Setting this distribution as $f(\alpha)$ and using the aforementioned algorithm, I derive some solutions for $v(\cdot)$, which are depicted in Fig. 3.3. There are some interesting things to note. First (and somehow expected), the $v(\cdot)$ function is not symmetric in zero. This comes from the fact the $f(\alpha)$ is not symmetric either, and therefore the only way for $v(\cdot)$ to match condition 3.11 is not to be symmetric[15]. Also, even if we can have different functions $v(\cdot)$ that meet this condition, both sides of it have a similar shape with respect to zero (although their ratio is not constant).[16] Following the remark that both sides of the solutions have a similar shape, it could be that restricting the solution of $v(\cdot)$ to be concave, departing from zero to both sides, could give us an unique solution, if the goal

[15] An easy way to see this is to check for the extreme agents $\alpha = 0$ and $\alpha = 1$.

[16] For the solution depicted in solid line, a guess function $v_o(j) = |j|$ was used. For the other two examples, signified by dashed and pointed lines, more concave curves (always symmetric in zero) were used. The asymmetry is reached by the algorithm itself.

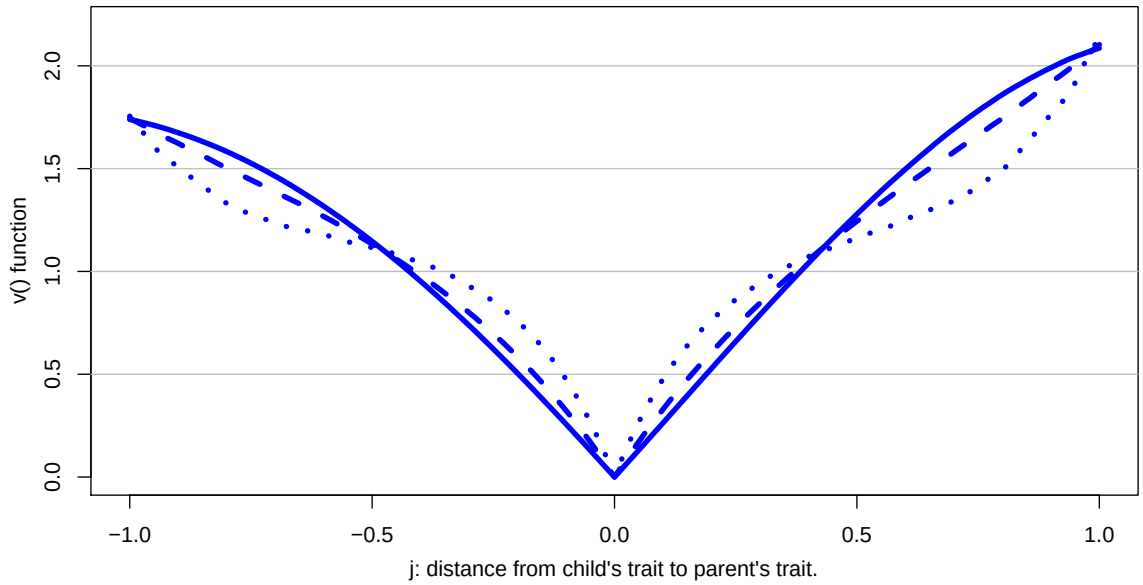


Figure 3.3: Examples of solutions for $v(\cdot)$.

were to find only one solution.

Second, an interesting feature to note is that this difference on both sides of $v(\cdot)$ means that a fully homo-oeconomicus parent has a greater 'dislike' (or disutility) of having a fully Kantian child than the way around. Linked to this observation and to the previous paragraph, we also find that the ratio of the two extremes of $v(\cdot)$, namely $v(-1)/v(1)$, is quite close to the ratio of the two clusters of people, $0.2981/0.3611$, no matter which solution of $v(\cdot)$ we use (for this given transformation between γ and α).[17] These two clusters, as was shown in Fig. 3.2b, agree with the semi-uniform part of the distribution of $f(\alpha)$.

Recalling that having population dynamics with direct and oblique transmission means that a parent will exert effort depending on the actual distribution of the population and the shape of the function $v(\cdot)$, we could think of this function as being a distillation of the parents dislike or disutility of having a child with a different trait compared to theirs. Therefore, this means that the result of having a population with a (slight) majority of homo-oeconomicus people comes, at least in part, from the fact that these people consider being Kantian to be a much worse option. One line of thought that could be explored is the following: it seems that homo-oeconomicus people are selfish compared to Kantian

[17] Using other transformations between γ and α, as in Appendix 3.E will change this ratio, and in that case this statement might not hold true any more. This might signal that the distribution shown in Fig. 3.2b would be a better fit.

ones, since the latter care for other people. Therefore, it seems reasonable to think that homo-oeconomicus people use a stronger myopic empathy when evaluating their offspring's utility, as compared to Kantian parents. Since the essence of the function $v(\cdot)$ comes from this myopic empathy, this explanation fits well the difference between the positive and negative side of the function $v(\cdot)$: They are the homo-oeconomicus dislike and the Kantian one, respectively. It would be interesting to better understand the social and psychological reasons behind this point, although this vein of thought escapes the scope of the present paper.

Another compelling reason to explore this last point more precisely would relate to understanding how to change a society that is not exhibiting green behaviour into a green one, in line with the topic discussed in Chapter 1. If there is a way to modify the function $v(\cdot)$ through education, active information or another social manner, then the society would become more Kantian, which would in turn make it greener.

3.4 Conclusions

The present work connects the trait transmission mechanism to the well-known results in Dictator Games experiments. To do so, first I develop a continuous version of a trait transmission process and I solve for the question: Which trait 'transition function' $v(\cdot)$ would make a population evolve into a specific distribution of that trait? To answer this question, I find the condition that has to be met and develop an algorithm for those cases where analytical solutions are not possible.

Then by mapping the results of the DG into a Kantian trait distribution, I rephrase previous question into the following one: What force is behind this observed distribution of a specific trait, that trait being in this case the Kantian morale? This force has to do with parents' desire, more or less, of having their offspring resemble themselves. It turns out, as was already clear in the trait transmission literature, that this force shapes the constitution of the society. In the specific case of a Kantian trait and using DG experiments' results, I find that homo-oeconomicus people have a stronger dislike or disutility of having a child with a different trait as compared to their Kantian counterparts. Following the origin of the parent's will behind the trait transmission, we know that myopic empathy is (at least one) reason for wanting our children to be as we are. Hence, homo-oeconomicus people seem to be more myopic than the Kantian ones, which make sense when we consider the definitions of being Kantian and homo-oeconomicus. Homo-oeconomicus people tend to be more selfish, where Kantian ones are more empathic. It turns to be ironic, assuming all these assumptions as true, that Kantian people, in caring more for their fellows, are jeopardizing their own (evolutionary) existence.

As mentioned in the previous section, it would be interesting to better understand the origins of the function $v(\cdot)$, the one responsible for the transmission forces that shape our society, at least in the Kantian arena. A better understanding is not only appealing for its own merits, but could also help to figure out how to move societies to become greener. A second line of future research could to find out a more general mathematical solution of the function $v(\cdot)$.

It could also be interesting to pursue the following vein: Having found different trait distributions for different countries (as with, for example, the Kantian trait distribution), one could examine if these differences are related to other results in those countries (for example, those concerning environmental conservation, other games played, or real green behaviour). If the previous idea turns out to be true, then one could investigate how to modify the $v(\cdot)$ function, the cost of transmitting a Kantian trait (the $H(\cdot)$ function), or something along those lines in order to induce and sustain greener societies.

One more fairly direct extension that could be pursued has to do with finding out this 'intrinsic' motivation for other types of traits. The Kantian trait discussed above can be related to how generous people are and how this generosity is transmitted from generation to generation. Other social traits could be analysed in the same fashion.

A final idea to explore involves relaxing the assumption that society is at equilibrium. First, we should verify if different cohorts behave (significantly) differently from one another. If they do not, the present findings would hold true. But if they do, we could explore how they change (speed and shape), where it might be possible to derive the $v(\cdot)$ function from, and possibly other interesting properties.

Bibliography

Alberto Bisin and Thierry Verdier. A model of cultural transmission, voting and political ideology. *European Journal of Political Economy*, 16(1):5–29, March 2000.

Alberto Bisin and Thierry Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298 – 319, 2001. ISSN 0022-0531.

Christoph Engel. Dictator games: a meta study. *Experimental Economics*, 14(4):583–610, 2011.

Daniel Friedman. On economic applications of evolutionary game theory. *Journal of Evolutionary Economics*, 8(1):15–43, 1998.

Marc Harper. Information geometry and evolutionary game theory. *arXiv preprint arXiv:0911.1383*, 2009.

Esther Hauk and Maria Saez-Marti. On the cultural transmission of corruption. *Journal of Economic Theory*, 107(2):311 – 335, 2002. ISSN 0022-0531.

Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, 2003.

Jean-Jacques Laffont. Macroeconomic constraints, economic efficiency and ethics: An introduction to kantian economics. *Economica*, 42(168):430–37, November 1975.

Julie Novakova and Jaroslav Flegr. How much is our fairness worth? the effect of raising stakes on offers by proposers and minimum acceptable offers in dictator and ultimatum games. *PLoS ONE*, 8(4), 2013.

Tanya O'Garra and Dave Krantz. Bystander effects and compassion fade: An experimental investigation of how giving varies with the number of dictators and recipients. *Working paper presented at 5th WCERE Conference (World Congress of Environmental and Resource Economics), Istanbul, Turkey (28 June-2 July 2014)*, 2014.

Maria Sáez-Martı and Yves Zenou. Cultural transmission and discrimination. *Journal of Urban Economics*, 72(2):137–146, 2012.

Karl Sigmund. A survey of replicator equations. In *Complexity, Language, and Life: Mathematical Approaches*, pages 88–104. Springer, 1986.

Gerald Silverberg. *Evolutionary modeling in economics: recent history and immediate prospects*. Citeseer, 1997.

E Christopher Zeeman. Population dynamics from game theory. In *Global theory of dynamical systems*, pages 471–497. Springer, 1980.

3.A General evolution equation for a n type population

Having an n-type population, let us call q_t^i the share of type i at time t. Recalling the definition of the transition probabilities given in Eqn. (3.5), we have:

$$P^{ii} = \tau^i + (1-\tau^i)q_t^i \qquad P^{ij} = (1-\tau^i)q_t^j \qquad \forall i \neq j \tag{A.1}$$

Therefore, the share of type i at time $t+1$ will simply be:

$$\begin{aligned}
q_{t+1}^i &= P^{ii}q_t^i + \sum_{j\neq i} P^{ji}q_t^j \\
&= \Big(\tau^i + (1-\tau^i)q_t^i\Big)q_t^i + \sum_{j\neq i}\Big((1-\tau^j)q_t^i\Big)q_t^j \\
&= q_t^i\Big(\tau^i + (1-\tau^i)q_t^i + \sum_{j\neq i}(1-\tau^j)q_t^j\Big) \\
&= q_t^i\Big(\tau^i + \sum_{j=1}^{n}(1-\tau^j)q_t^j\Big) \\
&= q_t^i\Big(\tau^i + \sum_{j=1}^{n} q_t^j - \sum_{j=1}^{n}\tau^j q_t^j\Big) \\
q_{t+1}^i &= q_t^i\Big(\tau^i + 1 - \overline{\tau}\Big) \\
\Delta q_t^i = q_{t+1}^i - q_t^i &= q_t^i\,(\tau^i - \overline{\tau}) \qquad\qquad \text{with} \qquad \overline{\tau} = \sum_{i=1}^{n}\tau^i q_t^i
\end{aligned}$$

3.B Solving for an specific $f(\alpha)$

I find some solutions $v(j)$ that solve problem defined by the Eqn. (3.11), rewritten here:

$$\int_0^1 v(j,\alpha)f(\alpha)d\alpha = \tau(j) = \text{constant} \qquad \forall j \mid f(j) > 0 \tag{B.2}$$

for $f(\alpha) = C_1\,\delta(0) + C_2 + C_1\,\delta(1)$ with $2C_1 + C_2 = 1$. One way to solve this integral equation is to make use of non-standard calculus and performing the computations within the *hyperreals* (approach that is more straightforward). With this set-up, we have that $f(0)$ (where the first Dirac delta is centred) is equal to a infinite hyperreal, such that $f(0)\cdot\epsilon = C_1$, with ϵ being an infinitesimal hyperreal. In the same fashion, we have

that $f(1) \cdot \epsilon = C_2$. Hence,

$$\begin{aligned}
\tau(j) = \int_0^1 v(j,\alpha) f(\alpha) d\alpha &= v(j)\, f(0)\, \epsilon \,+\, C_2 \int_\epsilon^j v(j-\alpha) d\alpha + C_2 \int_j^{1-\epsilon} v(j-\alpha) d\alpha \,+\, v(1-j)\, f(1)\, \epsilon \\
&= v(j)\; C_1 + C_2 \Big(\int_\epsilon^j v(j-\alpha) d\alpha + \int_j^{1-\epsilon} v(j-\alpha) d\alpha \Big) \,+\, v(1-j)\; C_1 \\
\text{(returning to the real domain)} &= C_1 \big(v(j) + v(1-j) \big) + C_2 \Big(\int_0^j v(j-\alpha) d\alpha + \int_j^1 v(j-\alpha) d\alpha \Big) \\
\text{(changing variable inside integrals)} &= C_1 \big(v(j) + v(1-j) \big) + C_2 \Big(\int_0^j v(x) dx + \int_0^{1-j} v(x) dx \Big) \\
\text{(replacing } w(j) = \int_0^j v(x) dx) &= C_1 \big(v(j) + v(1-j) \big) + C_2 \big(w(j) + w(1-j) \big) \qquad \text{(B.3)}
\end{aligned}$$

which has to be constant (equilibrium condition).

Finding general solutions for this problem can be a colossal task. One approach to find some solutions is to look for them in following equations:

$$C_1\, v(j) + C_2\, w(j) = \text{constant} \qquad \text{for } 0 \le j \le 1 \qquad \text{(B.4)}$$

$$C_1\, v(j) + C_2\, w(1-j) = \text{constant} \qquad \text{for } 0 \le j \le 1 \qquad \text{(B.5)}$$

It is easy to see that solutions for (B.4) and (B.5) are also solutions for (B.3). This approach obviously restricts the possible solutions to be found, but it also eases that task at hand.

Using (B.4), I found that the family of functions $v(j) = K - a \cdot c \cdot e^{-a|j|}$ (with $a = C_2/C_1$) is a solution for $f(\alpha) = C_1\, \delta(0) + C_2 + C_1\, \delta(1)$ with $2C_1 + C_2 = 1$. I verify this by checking that the following expression is constant:

$$\begin{aligned}
\tau(j) &= C_1 \big(v(j) + v(1-j) \big) + C_2 \big(w(j) + w(1-j) \big) \\
&= C_1 \big(v(j) + v(1-j) \big) + C_2 \Big(\int_0^j v(x) dx + \int_0^{1-j} v(x) dx \Big) \\
&= C_1 \big(K - ac\, e^{-a|j|} + K - ac\, e^{-a|1-j|} \big) + C_2 \Big(\int_0^j K - ac \cdot e^{-a|x|} dx + \int_0^{1-j} K - ac \cdot e^{-a|x|} dx \Big) \\
&= 2KC_1 - acC_1 e^{-aj} - acC_1 e^{-a(1-j)} + C_2 \Big(Kx \Big|_0^j + ce^{-ax} \Big|_0^j + Kx \Big|_0^{1-j} + ce^{-ax} \Big|_0^{1-j} \Big) \\
&= 2KC_1 - acC_1 e^{-aj} - acC_1 e^{-a(1-j)} + C_2 \Big(Kj + ce^{-aj} - c + K(1-j) + ce^{-a(1-j)} - c \Big) \\
&= 2KC_1 + C_2 K - 2C_2 c + ce^{-aj}(C_2 - C_1 a) + e^{-a(1-j)}(C_2 - C_1 a) \\
&= 2KC_1 + C_2 K - 2C_2 c \qquad \text{(since } a = C_2/C_1) \\
&= \text{constant}
\end{aligned}$$

In a similar way, it can be shown that $v(j) = b \cdot sin\left(\frac{(4k-3)\pi}{2} \cdot |j|\right)$ is a solution. In this case, we focus on solving for (B.5). Because of the properties of the trigonometric functions, in particular their derivatives and reflections in $\pi/2$,[18] we are able to find these types of solutions. The idea is to make the range of $[0, 1]$ of j coincide, with a change of variable, with $[0, \pi/2]$, $[0, 5\pi/2]$, etc. The proof follows the same line as the one before and it is left to the reader.

3.C Properties of the Algorithm

As stated in Section 3.3, I show that if at some iteration k, $v_k(j)$ is a solution, the algorithm converges, and that if it converges then $v_k(j)$ is a solution.

Let $v_k(j)$ be a solution. This means that $\tau_k(j)$ is constant for $0 \leq j \leq 1$. Therefore, the term $1/\tau_k(j)$ is also constant, and hence the adjustment function $a_k(j) = \frac{1}{\tau_k(1-2j)}$ is also constant. Let call this last constant K. From the iteration process we see that:

$$v_{k+1}(j) = a_k(j) \cdot v_k(j) = K \cdot v_k(j)$$

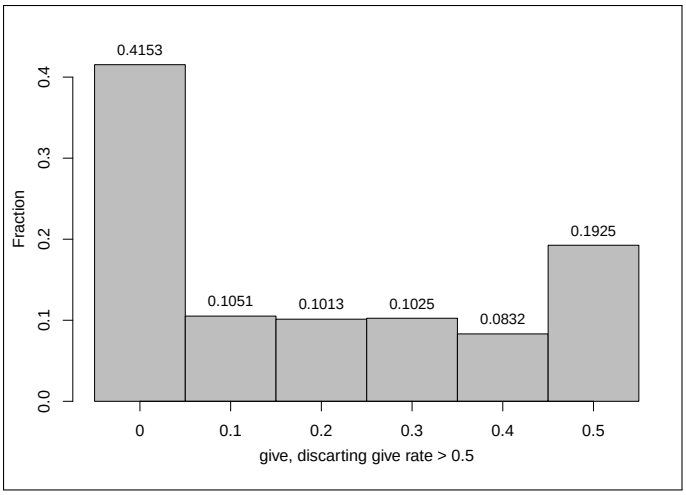
It is easy to see that $K = 1$. If not, having $K < 1$, would yield to $v_k(j) = 0$, when $k \to \infty$, which would mean $\tau_\infty(j) = 0$ and then $a_\infty(j) = \infty$, which is a contradiction. Similarly, if $K > 1$ we get that $a_\infty(j) = 0$, again a contradiction. Therefore, $K = 1$.

On the other hand, if the iteration converges, it means that $v_{k+1}(j) = v_k(j)$ for k bigger than some fixed number. This means that $a_k(j) = 1$, which in turn means that $\tau_k(j) =$ constant, proving that $v_k(j)$ is a solution.

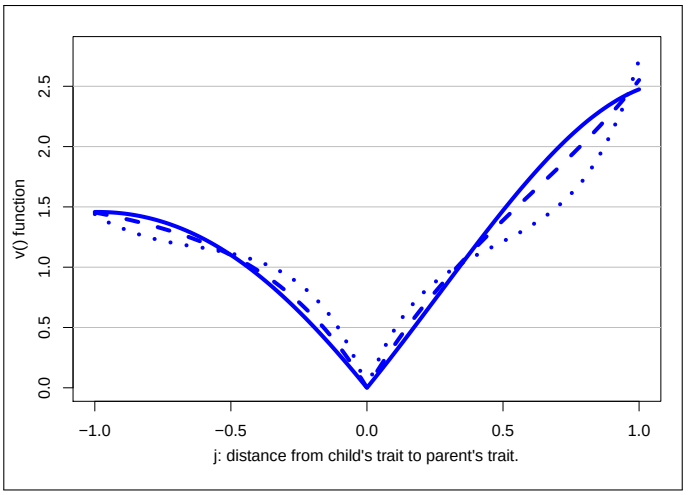
3.D Alternative Case: Disregarding give rates over 0.5

Here I find the $v(\cdot)$ solution for the Kantian trait of the DG, using a truncated version of Engel's (2011) data. As stated in Section 3.3, here I perform the same operations, but I discard all give rates bigger than 50%. To do this, I drop this information and recalculate the fractions of give rates within this new universe. I then transform this information into its Kantian distribution equivalent, and finally I find the solution function $v(\cdot)$ using the algorithm previously mentioned. The histogram of the truncated information (give rate) and some solutions are plotted in the following figure. To produce these solutions, the same initial guesses for $v_0(\cdot)$ were used as in Section 3.3. As we can observe, the solution $v(\cdot)$ is once again asymmetrical, although its 'sides' are yet alike. The difference between this and the function depicted in Fig. 3.3 (Page 114) is that in this one, the two sides ($v(+)$

[18] Meaning, for example, that $sin(\pi/2 - \theta) = cos(\theta)$.



(a) Truncated distribution of give rates, using Engel (2011).



(b) Solutions of $v(\cdot)$ for the corresponding Kantian distribution.

Figure 3.a: Truncated data for give rates and some solutions of $v(\cdot)$.

and $v(-))$ are more different, following the stronger dissimilarity of the new clusters in $\alpha = 0$ and $\alpha = 1$ (Fig. 3.aa). However, the essential result holds: the homo-oeconomicus agent has a stronger dislike of having a Kantian child than the opposite case.

3.E Using different utility functions for transforming give rate into Kantian trait.

As stated in Section 3.3, choosing different utility functions in Eqn. (3.23) produces different transformations from give rate γ into Kantian trait α. Always using a CRRA function, I depict different transforms for different values of ϵ, their risk aversion measure:

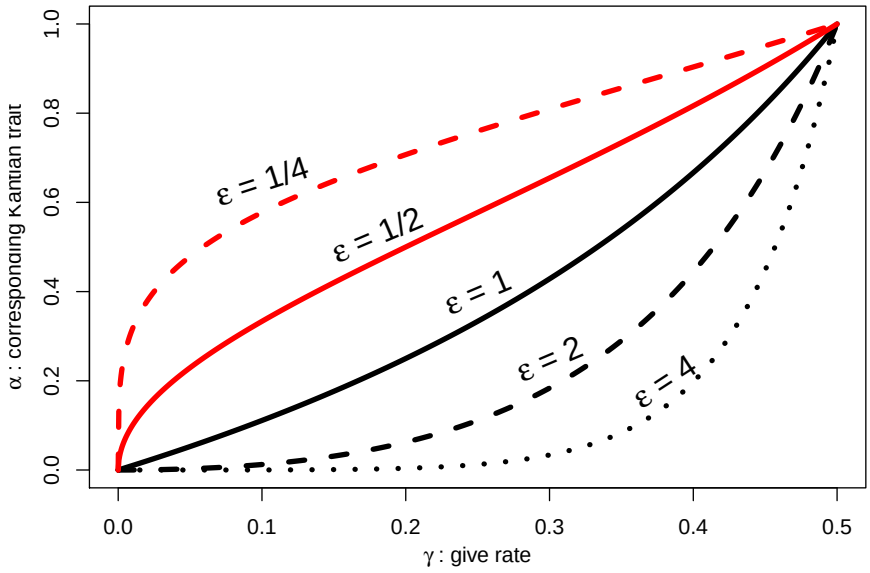
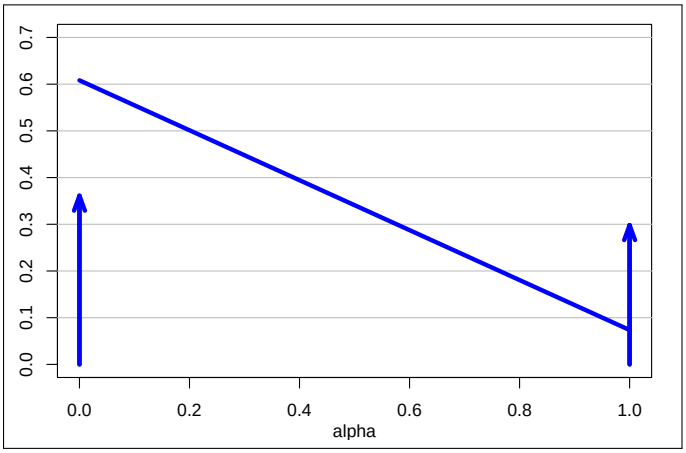


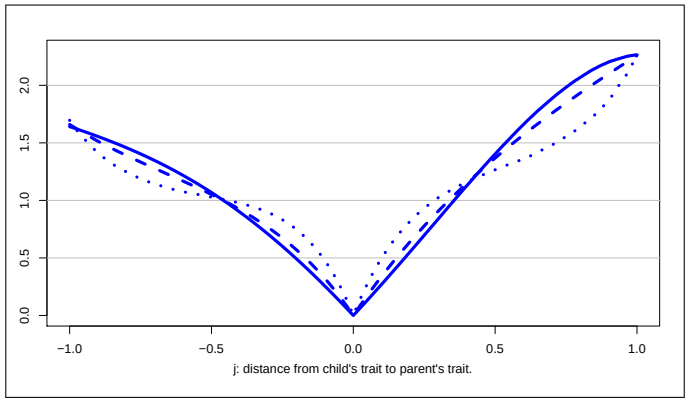
Figure 3.b: Examples of transforms from give rate to Kantian trait.

As one can observe from Fig. 3.b, using measures of risk aversion too different from $\epsilon = 1$ produces a loss of information. For example, with $\epsilon = 4$, values of give rate between zero and 0.25 yield to almost the same Kantian trait, around zero. We find a similar result if we use bigger values, as in the case of $\epsilon = 1/4$.

I turn now into how the distribution of the Kantian trait α could be, depending on the choice of ϵ and then, how this would translate info the solution of $v(\cdot)$. In the following figures I plot the case with $\epsilon = 0.9$ and $\epsilon = 1$:

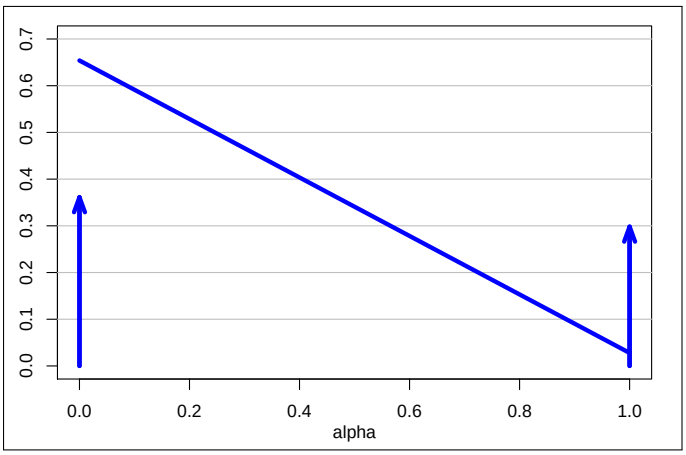


(a) Distribution in Kantian measurement with $\epsilon = 0.9$

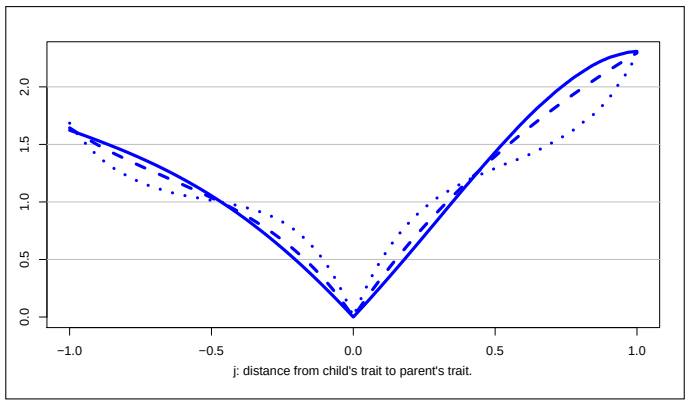


(b) Solution $v(\cdot)$, with $\epsilon = 0.9$

Figure 3.c: Results using $\epsilon = 0.9$



(a) Distribution in Kantian measurement with $\epsilon = 1$



(b) Solution $v(\cdot)$, with $\epsilon = 1$

Figure 3.d: Results using $\epsilon = 1$