11. Tipping and reference points in climate change games

Alessandro Tavoni and Doruk İriş

1. INTRODUCTION

We live in a world characterized by discontinuities, where thresholds for abrupt and irreversible change are omnipresent, both in economic and ecological dynamics. Such thresholds, often referred to as tipping points, trigger nonlinear responses on the part of individuals or ecosystems.

Climate change is a prominent example of the pervasiveness of tipping points, since they appear both in the strategic decision to embark in costly mitigation (Heal and Kunreuther, 2012) and in the Earth's climate system (Lenton et al., 2008).

In this chapter we will focus on "behavioral tipping points", to distinguish them from the ecological ones. As it will become apparent, though, the two are closely linked, since planetary boundaries define "the safe operating space for humanity with respect to the Earth system and are associated with the planet's biophysical subsystems or processes" (Rockström et al., 2009). Hence, to discuss strategies one has to account for the underlying physical processes and how they are perceived (Tavoni and Levin, 2014).

Whether a country or a subnational actor (a city, an NGO or a firm) decides to invest in a clean technology, or more broadly in actions aimed at reducing greenhouse gas emissions, depends on its expectations with regards to the actions of others. This is particularly salient in the context of a public good such as climate change mitigation. Depending on the choice of the underlying parameters, the climate change game is generally characterized either by coordination (selecting the mutually preferable outcome in a chicken game), or a unique inefficient outcome resulting from widespread defection in a prisoner's dilemma game (Barrett, 2016). In the latter class of games, which comprises linear public goods games for plausible parameters capturing the temptation to defect (not contributing to the mitigation good), the worst-case scenario is for an actor to take costly action while the others refuse to do so, the so-called "sucker's payoff". Arguably, this may have been the case for the European Union in climate negotiations up to COP 20 in Lima, with unilateral commitments by the EU routinely unmatched by other large economies. COP 21 in Paris was perhaps the first Conference of the Parties to mark a greater willingness to show leadership by other large powers, such as China and the United States (although the election of Trump as president casts a long shadow over the prospects of the Paris agreement). A possible interpretation, in keeping with the above arguments, is that enough action at various scales had accumulated in the years leading to the Paris summit that even less committed countries showed an increased willingness to act.[1]

[1] Indeed, while the Nationally Determined Contributions agreed upon in Paris are insufficient to meet the target of "holding the increase in the global average temperature to well below 2°C

These emerging trends are potentially game-changing, provided that enough actors lead the way by taking action early on. Once a tipping point for sufficient investments in low carbon technologies has been reached, and constituencies with stakes in the nascent markets have formed, standard economic forces will sustain the transition to a carbon-neutral economy.

We will review some of the recent literature that provides clues about when such reinforcing dynamics take place. In doing so, we will come across related concepts, such as diffusion and feedback. Importantly, given the wide scientific uncertainties surrounding the location of the thresholds, we will discuss the role of expectations and argue that reference points are crucial for supporting cooperation. Intuitively, under uncertainty asymmetries about views on the expected losses from climate change are as important as differences in objective (but elusive) vulnerabilities.

2. REGIME SHIFTS AND CATASTROPHIC CLIMATE CHANGE

A key concept relating to tipping points, or regime shifts as they are otherwise referred to, is that of irreversibility: "Ecological regime shifts are large, sudden changes in ecosystems [. . .]. They entail changes in the internal dynamics and feedbacks of an ecosystem that often prevent it from returning to a previous regime" (Biggs et al., 2009). In other words, once a threshold has been passed, the system will enter a new basin of attraction, resulting in a sudden and persistent shift to a new stable fixed point (fixed point 2 in Figure 11.1).

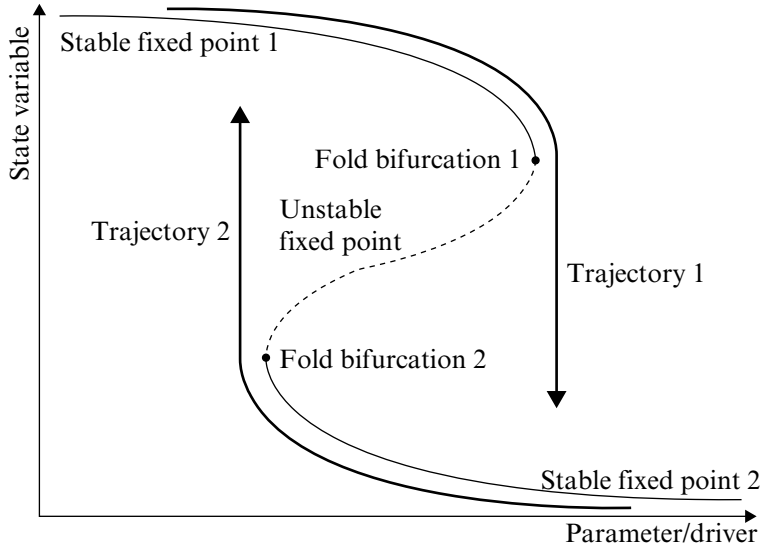
A related concept in the social sciences is that of diffusion, which Rogers (1995) defines as "the process by which an innovation is communicated through certain channels over time among the members of a social system". The link here is the idea that the actions of others can reinforce one's own choices. Different terms have been coined for this feedback mechanism, depending on the disciplinary focus, such as bandwagon effect in fashion-oriented behavior (Leibenstein, 1950), adoption thresholds (Granovetter, 1978), entrapment (Dixit, 2003), global cascades (Watts, 2002) and tipping (Gladwell, 2000), to mention a few.

Early models of diffusion focus on the societal adoption rate (of a technology or behavior), whose dynamics are governed by the overall ratio of adopters to non-adopters at a given point in time (Bass, 1969; Young, 2009). The main insight from the Bass model, also known as S-shaped diffusion curve, is that diffusion follows a nonlinear trend, with a fast acceleration in the initial phase of adoption and a subsequent saturation (Figure 11.2).

Of course, the channels through which innovations diffuse are less mechanistic than those depicted in Figure 11.2. An important driver of the speed of diffusion, whether of an opinion, a fad or a technology, is the topology of the network on which the agents are embedded, since it will spread through society according to dynamics that depend on the patterns of social connections (Currarini et al., 2015).

---

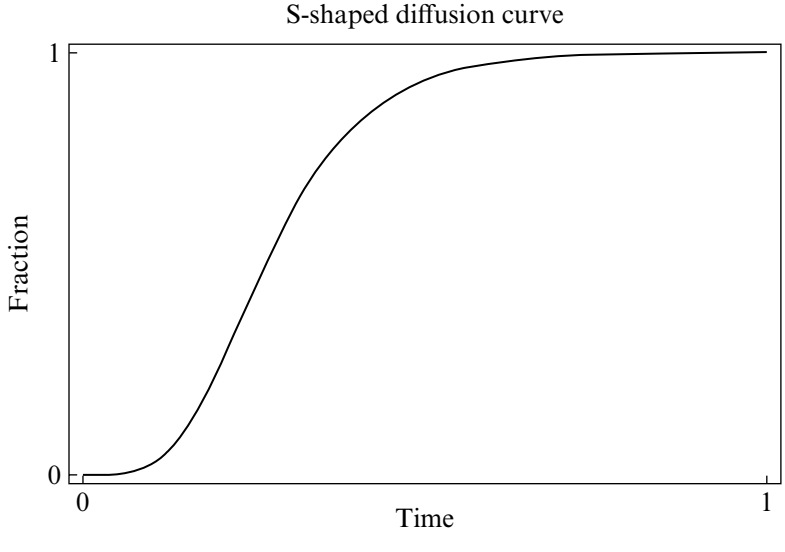
above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5°C above preindustrial levels", these commitments are viewed as genuine, in the sense that they exceed what states would have done in the absence of the agreement (Averchenkova and Bassi, 2016).



Note: Fold bifurcations occur when a stable fixed point (solid lines) collides with an unstable fixed point (dashed line), and can lead to regime shifts.

Source: Lade et al. (2013).

Figure 11.1 Regime shifts



Note: The fraction of adopters is plotted against time.

Figure 11.2 An example of S-shaped diffusion curve

What are the implications of the nonlinearities identified above in ecological and social processes for the economic analysis of climate change? We hope to contribute to answering this question by reviewing some of the recent literature focusing on the avoidance of a threshold for dangerous climate change.

Perhaps the most influential paper on the economics of catastrophic climate change is due to Weitzman (2009), who focused on high-impact, low-probability catastrophes and catalyzed attention on the so-called (bad) fat tail of the probability density function (PDF) of what might happen in the absence of serious mitigation policy.[2] His dismal theorem points to the "potentially unlimited exposure to catastrophic impacts" in the presence of "deep structural uncertainty in the science coupled with an economic inability to evaluate meaningfully the catastrophic losses from disastrous temperature changes" (Weitzman, 2009). The implications for both theorists and practitioners are big: if we are serious about accounting for unlikely, yet possible future climate disasters, we cannot rely on the standard cost–benefit analysis calculus dealing with uncertainty in the form of a known PDF with thin tails. Consequently, Weitzman advocates a "generalized precautionary principle" that accounts for the potential catastrophic costs of inaction.

A not-so-dismal account is found in Heal and Kunreuther (2012), which instead investigates the implications of the existence of a tipping point in the adoption of climate policies by the international community. The authors offer illustrative evidence on the role of early adopters (in the abstract representation of Figure 11.2, those who are located at the bottom of the diffusion curve) in triggering a global shift away from the use of damaging pollutants, towards greener alternatives. They cite two pieces of evidence. One concerns the adoption of unleaded gasoline in replacement of leaded gasoline: the unilateral adoption by the United States meant that the subsequent adoption's costs for other countries was limited to modifying refinery capacity, since motor industries exporting to the United States had to transition to lead-free fuel immediately after the move. Thanks to these reduced costs for the followers, the new technology spread quickly worldwide. The second example refers to phasing out chlorofluorocarbons (CFCs), a remarkable achievement of the Montreal Protocol on Substances that Deplete the Ozone Layer. In this case, the U.S. decision to sign the Montreal Protocol hinged on a technological innovation by Du Pont, the world's largest producer of CFCs, allowing the company to gain from elimination of CFCs. As in the previous example, strategic complementarity led most countries to phase out ozone-depleting chemicals.

3. KEY FEATURES OF DANGEROUS CLIMATE CHANGE

Avoiding disastrous climate change requires overcoming several barriers to collective action. As other public goods situations, such as protecting the global climate from *gradual* climate change, *dangerous* climate change requires widespread cooperation in the face of individual incentives to free-ride on the efforts of others. Moreover, it is characterized by sudden transitions to harmful states (tipping points) and irreducible uncertainty

[2] A fat-tailed distribution assigns a higher probability to rare events in the extreme tails than does a thin-tailed one.

on the location of the threshold and on the consequences of crossing it (Dannenberg and Tavoni, 2016). A complicating factor is that the ability to contribute to the solution and benefits from doing so vary greatly among parties. Together, these characteristics compound the difficulty of securing sufficient mitigation effort, with potentially disastrous consequences.

The Framework Convention on Climate Change has warned that climate change may be "abrupt and catastrophic", rather than linear and smooth, once greenhouse gas concentrations in the atmosphere exceed a certain threshold. This threshold is commonly identified as the concentration level that would translate in a 2°C average temperature increase above preindustrial levels. More recently, the Paris Agreement raised the ambition to limiting the increase to 1.5°C. The scientific literature confirms the possibility of dangerous thresholds but also shows that there is large uncertainty about their location (Rockström et al., 2009). This uncertainty is compounded by the political and technical uncertainty arising from translating such thresholds into the necessary mitigation measures required to avert disaster (on top of the uncertainty about the economic value of the catastrophic losses from failure to avoid dangerous climate change). In addition, international climate negotiations that aim to reduce global greenhouse gas emissions are strongly influenced by a conflict between rich and poor countries. Consequently, the Kyoto Protocol addressed this North–South equity issue by recognizing the industrialized nations' special responsibilities through the principle of "common but differentiated responsibilities". However, major polluters and great powers, such as China and the U.S., have until recently proven reluctant to bind themselves to internationally agreed ambitious emission reductions. While progress has been made in Paris, the move to a bottom-up architecture where states unilaterally announce their proposed targets confirms the burden-sharing difficulties.

How can this deadlock be broken? As real-world data on alternative mechanisms is not available, we must rely on theory and experimental economics. Here, we focus on several simple and recent applications of both methods. Although both theory and experiments rely on simplified settings, and in the case of experiments on convenience samples, these methods have the advantage of allowing one to isolate in a controlled manner the effect of relevant features (such as inequality and uncertainty) on cooperation. The next two sections tackle, in turn, two classes of games that have recently been modified to accommodate some of the features we have discussed thus far in the context of the avoidance of disastrous climate change: public goods games and coalition formation games.

4. THRESHOLD PUBLIC GOODS GAMES

In order to account for the threat of disastrous losses from abrupt and irreversible climate change, a recent literature has resorted to studying public goods games, both theoretically and in the laboratory. These can accommodate the above features by introducing a discontinuity, in the form of a threshold for dangerous climate change. Threshold public goods games (TPGs), unlike linear ones, feature multiple Nash equilibria. Restricting attention to symmetric play, we have two equilibria: a Pareto-superior one where players shoulder the same burden and avoid catastrophe *just*, alongside the unique (Pareto-inferior) equilibrium prescribing defection. The latter characterizes the quintessential cooperation

problem of under-provision (or lack of provision) of public goods. Thus, the presence of a discontinuity in the form of a threshold beyond which losses shoot up appears to facilitate the problem by effectively serving as a coordination device (Barrett, 2013).

In addition to work that focuses on equilibrium behavior as in traditional game theory, variants of the above TPGs have also been employed in evolutionary game theory models (EGT) to investigate off-equilibrium dynamics. The premise behind the replicator dynamics, the most commonly employed imitation dynamics in EGT, is that agents imitate the behavior of the other players in the population whenever these appear to be more successful. While the implied myopic behavior represents a departure from traditional game theory, which models optimization through instantaneous best responses of all players to one another, long-run behavior in evolutionary games often coincides with static Nash equilibria (Weibull, 1997).

Vasconcelos et al. (2014) investigate the collective action problem posed by disastrous climate change using EGT to assess disaster avoidance success in TPGs. Among other features, they focus on two that are relevant for our discussion: inequality and polycentric governance (Ostrom, 2009). To tackle the former, they include an asymmetric distribution of wealth that mimics the existing inequalities among nations. The authors find that the combination of homophily (high likelihood of imitating the strategies of agents similar to oneself) and inequality can lead to the collapse of cooperation, due to a drop in the poor countries' contributions. To assess the effectiveness of local sanctioning institutions in coping with dangerous climate change, Vanconcelos and colleagues distinguish between a global setup, in which the entire population constitutes the sole negotiating group, and a local setup, in which individuals interact in several smaller groups. They find that local institutions are more successful at promoting the emergence of widespread cooperation, particularly under low risk perception (when the perceived probability of disaster under non-provision of the public good is small). The reason is that cooperation can only thrive if at least some cooperators in the population do better than the defectors, and this is more likely to happen in the local setup because it allows for more heterogeneity in behavior.[3]

In sum, the above findings cast an optimistic light on the prospects for climate negotiations, especially in the aftermath of the Paris Agreement, which was hailed as an unprecedented success and relied heavily on unilateral pledges, known as Intended Nationally Determined Contributions (INDCs). However, as noted above, uncertainty about the location of the threshold complicates matters, by bringing us back to a gradual losses setup, as the sharp discontinuity at the threshold is replaced by a smooth probability density function of expected losses. Experiments confirm the detrimental effect of threshold uncertainty; see Barrett and Dannenberg (2012) and Dannenberg et al. (2014). A crucial question is then to what extent can climate treaties coordinate actions in spite of structural uncertainty. To address it, we turn to the literature on the economics of international environmental agreements, before returning in greater detail to the experimental evidence.

---

[3] Note however that the local institutions are assumed to be independent of each other. Hence, further research is needed to clarify if the results also hold in a more realistic setting with spillovers across regions.

5. COALITION FORMATION GAMES

Building on the work of d'Aspremont et al. (1983), international environmental agreements have long been modeled in game theory through coalition formation games (Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 2016). The premise is that the equilibrium number of signatories to a self-enforcing international agreement follows from the conditions of internal and external stability, which respectively guarantee that no signatory is better off leaving the coalition, and that there is no incentive for a non-signatory to join the coalition. These conditions are required since treaties such as the Kyoto Protocol (let alone the Paris Agreement, which lacks legal force) cannot be enforced by external institutions and must therefore rely on incentives to overcome the compliance issue.

The accepted insight from the theory of international environmental agreements is that self-enforcing treaties fail to deliver, especially when cooperation is most needed (i.e. when the potential gains from cooperation are large due to high mitigation costs and high benefits from mitigation). Scott Barrett summarizes how the introduction of a known tipping point changes this calculus:

> The standard model of a self-enforcing international environmental agreement predicts that collective action in reducing greenhouse gas emissions will be grossly inadequate. When this model is modified to incorporate a certain threshold with catastrophic damages, treaties can become highly effective. If the benefits of avoiding the threshold are high relative to the costs, the prospect of catastrophe transforms treaties into coordination devices. (Barrett, 2013)

However, as discussed above, uncertainty is detrimental to the prospects of disaster avoidance:

> While the uncertain prospect of approaching catastrophes may commend substantially greater abatement in the full cooperative outcome, it may make little difference to non-cooperative behavior or to the ability of a climate treaty to sustain substantial cuts in global emissions. (Barrett, 2013)

What else can we then rely upon, to design an effective treaty? Is there a way out of the grim results found in most of the models reviewed so far? The game-theoretic literature on International Environmental Agreements (IEAs) has identified several mechanisms that have the potential to mitigate the issue of shallowness of the mitigation efforts (or small stable coalition size, both of which translate into unambitious treaties and increased threat of catastrophe).[4] These range from expanding the strategy space via side payments and issue linkage (Barrett, 2005), to introducing minimum participation rules and heterogeneity (Weikard et al., 2014) and imposing trade sanctions on non-participants (Nordhaus, 2015).

Preferences also matter for the outcome of IEAs. The literature has traditionally modeled negotiators as rational agents with standard preferences. A recent strand has instead explored the implications of departing from these assumptions. Examples include introducing preferences for equity (Lange and Vogt, 2003), for reciprocity (Nyborg, 2015), reference dependence (İriş and Tavoni, 2018), and appetite for campaign contributions by

[4] For a recent review, see de Zeeuw (2015).

policy makers subject to lobbying pressure (Marchiori et al., 2017).[5] We review some of the non-standard preferences literature (employing other regarding and reference-dependent preferences) in Section 7, where we restrict attention to coalition formation and public goods games (Bosetti et al., 2017).

Broadly speaking, all of the above modifications of the standard model ease the collective action problem, to some extent. The question remains, however, about which representation of the climate negotiations is more realistic. For instance, some improving mechanisms, such as imposing trade sanctions on those outside the "climate club", while theoretically desirable may prove difficult to implement due to the threat of retaliations and potential escalation to trade wars. Equally important, we lack empirical evidence about the relative effectiveness of different schemes, or about the preferences of the negotiators. We tackle these issues in the next section.

6. CLIMATE CHANGE EXPERIMENTS

A productive way to gather some relevant insights is to conduct controlled experiments with subjects who are assigned to different treatments aimed at capturing relevant features of the game. An obvious advantage is that one doesn't need to wait decades to assess how a given IEA, such as the one agreed in Paris in December 2015, has affected global emissions (in addition to the fact that one can easily compare treatments to the baseline "untreated" status quo in an experiment, while in reality the counterfactual is not easy to identify). The price to pay in order to have such control (internal validity) is that one has to greatly simplify the problem in the laboratory, possibly at a cost in terms of external validity.

In this section we briefly review some of the recent experimental literature that uses TPGs as a metaphor for studying cooperation on the avoidance of dangerous climate change. In these experiments small groups (of four to ten players) have the option to contribute part of their endowment to avoid a collective loss (in one or several successive rounds). The aggregate contributions are then evaluated to see how they compare to the investment required to avoid the tipping point, and a large fraction of the remaining endowment is lost if the target has not been met.[6]

Milinski et al. (2008) test the role of risk perception on the success rate in disaster avoidance by manipulating the probability of losing one's savings (p) if the group fails to invest the target sum (€120) by the end of ten successive rounds. Thus, we have a TPG with three treatments, corresponding to either 10%, 50% or 90% probability of loss when missing the provision threshold. They find that the ensuing share of groups who managed to avoid the loss are respectively 0%, 10% and 50%. One should not be surprised that none of the groups averted disaster when p=10%, since the expected loss is so small that it is collectively rational *not* to provide the public good (i.e., there is only mild climate change, and no collective action problem). It is instead noteworthy that even at very high expected costs from miscoordination from the Pareto-superior equilibrium of collecting €120 when

[5] For a paper focusing on reciprocal strategies in IEAs, see Ochea and de Zeeuw (2015).

[6] For a more comprehensive review of dangerous climate change games, see Dannenberg and Tavoni (2016).

p=90%, half of the groups fail the target (while still contributing approximately €113). One likely explanation for such spectacular coordination failure is that the subjects in this experiment were not allowed to signal their intentions or communicate in any way. As mentioned above, TPGs with certain threshold location is a coordination game, which naturally begs for communication opportunities.

Even in the face of inequality in the ability to contribute to the public good of loss avoidance, a clear coordination target proves a powerful mechanism to select the "good" equilibrium, provided that the communication channels are in place to facilitate redistribution. For instance, in a similar setting of repeated contributions to a TPG with p=50%, Tavoni et al. (2011) find that the majority of groups are able to avoid disaster when they have an opportunity to signal future contribution intentions via pledges. Namely, when communication was possible 60% of the groups with unequally wealthy participants (three "Poor" and three "Rich" countries) coordinated on disaster avoidance, compared to only 20% of successful groups when pledging was not an option. This is noteworthy, given the non-binding nature of pledges; yet, one should not be surprised that for a coordination game to be played well, communication is indispensable.[7]

Dannenberg et al. (2014) revisit this TPG, with groups of six players facing dangerous climate change with p=90%, by introducing uncertainty on the location of the tipping point. In two treatments they test the effect of risk, i.e., known (uniform) distribution over a range of potential thresholds, and ambiguity, i.e., lack of knowledge even on the distribution. They find that contributions become more erratic under threshold uncertainty, particularly so under ambiguity. Early commitment in the risk treatment, demonstrated by a willingness to invest in the public good early on and fulfill the pledges, helps groups to reduce the negative effect of uncertainty. This result resonates with the one by Tavoni et al. (2011) that the negative effect of wealth inequality is mitigated by leadership and communication.

In a one-shot TPG without sequential decision-making, Barrett and Dannenberg (2012) test a simplified version of the model developed in Barrett (2013), to assess the effect of uncertainty on two variables: the extent of the damages from dangerous climate change, and the location of the threshold. They find that while the first is unimportant for the prospects of catastrophe avoidance, when there is significant uncertainty on the location of the tipping point, the game reverts to a prisoner's dilemma and cooperation drops, with no group succeeding in avoiding the large loss.

Recently, scholars have begun to introduce political economy features, such as delegation, into catastrophe avoidance games mimicking dangerous climate change. The reason is simple: since the goal is to shed light on the ability of stylized climate agreements to improve upon non-cooperative behavior, one should aim to capture realistic aspects of their decision-making process. An intuitively relevant one is delegation: do elected delegates, representing the subgroup (country) to which they belong, by deciding how much to contribute to the public account on their behalf, behave more

[7] It appears that the positive effect of communication is indeed stronger than the negative effect of inequality. In the symmetric wealth treatments where endowments were equal across the six players, success rate in disaster avoidance went up less markedly when communication was allowed, from 50% to 70%. This is also to be expected, given that coordination is a much easier task under symmetric payoffs.

or less cooperatively than when acting independently? In other words, are the findings from the experiments mentioned above robust to delegation? İriş et al. (2019) compare the baseline case of four-subjects' groups where each "country" independently decides how much to contribute to the public good (again in the face of uncertainty), with two treatments where the countries are no longer singletons, but are themselves composed of a three-player constituency responsible to elect a delegate to represent them in the negotiations. Hence, while the number of countries is still four, in these two treatments the group is made of 12 subjects, given that in each country there is one delegate deciding for herself as well as for the two unelected candidates. The only difference between the two delegation treatments is that in one the delegate decides on contributions in the TPG without being exposed to public scrutiny from the constituency, while in the other the two non-delegates in each country are in the same room as their delegate and send their non-binding preferred contribution suggestions to the delegate. İriş et al. (2019) find that delegation without public pressure does not affect contributions much, relative to the baseline. However, even if messages are payoff-irrelevant in the experiment, public pressure has a significant negative effect on the delegates' contributions. This happens since the majority of delegates were elected because they signaled a low propensity to contribute in the practice phase, and, once elected, delegates focused on the lower of the two contributions preferred by their teammates, thus behaving more selfishly than in the treatments without public pressure. This finding echoes the one of Milinski et al. (2016), who use a similar setup with six three-player "countries" voting twice whether to confirm or vote out the incumbent representative. They find that selfish representatives are preferentially elected, and that once in power, they indeed contribute less to the public good than their fair share.

To summarize, in this section we have established the following: (i) the higher the (perceived) probability of disaster, the more likely it is that the catastrophic tipping point is avoided. (ii) Communication, in the form of pledges for achieving the target (e.g., INDCs), also increases the prospects for success, given that they facilitate the coordination problem; (iii) this is particularly needed in the presence of threshold uncertainty, although when uncertainty is too large failure is widespread. (iv) Delegates tend to focus on the least ambitious suggestion when confronted with public pressure, and act in a more self-interested manner than individuals.

In the next section we turn to the insights gathered in recent theoretical work on tipping points and reference dependence in related climate change games.

7. NON-STANDARD PREFERENCES

During the last decades, abundant evidence from both the laboratory and the field demonstrate that people exhibit discontinuities in their preferences as well. They often not only care about the outcome, but also about how it stands relative to a reference level. Economists and other social scientists argue that salient reference levels can relate to the status quo (Tversky and Kahneman, 1991), or alternatively emerge from social comparisons (Fehr and Schmidt, 1999; Rabin, 1993; Shafir et al., 1997), or can be based on goals and aspirations (Heath et al., 1999).

In this section, we focus on two widely used non-standard preferences: i) reference-

dependent preferences; and ii) other-regarding preferences (also referred to as social preferences). In particular, we review their applications in public goods and coalition formation games.[8]

7.1 Fairness and Other-Regarding Preferences

Standard economic models assume that economic agents are self-interested, meaning that they care only about their payoffs and are not concerned with others' payoffs. However, theoretical and empirical studies have shown that people have strong preferences for equity, even in market settings (Kahneman et al., 1986; Fehr et al., 1993; Fehr et al., 1997).

The literature classifies two groups of models for other-regarding preferences. The first group contains models in which people reciprocate based on the actions and perceived intentions of other players. The second group contains models in which people care about the distributions of payoffs.

If another player's action, or the outcome it leads to, is more positive (or negative) than the one that is considered fair—the reference level—then reciprocal players perceive it as a kind (or unkind) action. In the context of climate change, reciprocal players would be willing to undertake costly mitigation effort (e.g., abatement) insofar as others do too. Thus, reciprocal players do not only care about the outcomes of their actions, but also about the actions or intentions of other players (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Cox et al., 2007; Segal and Sobel, 2007).

Hadjiyiannis et al. (2012) is the first paper that incorporates reciprocity preferences into self-enforcing IEAs, in which reciprocal countries play an infinitely repeated prisoner's dilemma game. The authors compare the depth of the cooperation in the case of full participation in two games: one with reciprocal countries and the other with self-interested countries. The authors find that reciprocal countries that have moderate expectations with respect to others' national abatement strategies can support a greater degree of environmental cooperation than self-interested ones. They define reciprocal countries as those where abatement standards are considered fair if they are within certain boundaries. For moderate expectations, reciprocal countries perceive others' cooperative abatement levels as kind and non-cooperative (Nash) abatements as unkind, leading to more cooperation. However, when only very high abatement standards are deemed fair (i.e., fair abatement levels are higher than the most cooperative abatement levels countries can sustain), then reciprocity could have a detrimental effect on international environmental cooperation as countries perceive even the cooperative abatements as unkind. Their model therefore provides a novel perspective on the role of expectations in environmental negotiations. Consistent with their findings, failure in Copenhagen (COP15 summit), owing to the very high expectations, has been noted by some experts.[9] Consequently, leaders emphasized the

---

[8] DellaVigna (2009) reviews the empirical evidence of such non-standard preferences as well as other concepts in non-standard beliefs and non-standard decision making. Camerer et al. (2003) and Kahneman and Tversky (2000) collect early theoretical and applied studies pioneered behavioral economics. Camerer's (2003) and Kagel and Roth (Volume 1, 1995; Volume 2, forthcoming) are the three handbooks in experimental games.

[9] "Can Copenhagen still be saved?" *The Economist*, November 17, 2009.

importance of moderating expectations before and during the Paris conference (COP21) in December 2015.[10]

Nyborg (2015) studies the impact of reciprocal preferences on stable IEAs by using coalition formation games. To keep the model tractable, she focuses on a discrete strategy space, so that countries can either abate or pollute. Unlike Hadjiyiannis et al. (2012), she follows more closely Rabin (1993) in both incorporating reciprocity into the utility function and defining equitable payoff. She shows that no country participates in the coalition in the game with only self-interested countries. However, if some countries exhibit reciprocity, three stable coalition sizes become feasible: no participation, a minority coalition, and a majority or even full participation. The minority coalition improves upon zero participation, but only weakly. Moreover, for the majority or full participation coalition to be stable, most of the countries should exhibit strong reciprocal preferences.

A different agreement literature also studies the impact of reciprocity in which signatory countries *bind* themselves to share the cost of public good investment (Jang et al., 2016). Their model of reciprocity follows Dufwenberg and Kirchsteiger (2004), which extends Rabin's (1993) reciprocity model for a one-shot game to a sequential game with updating beliefs. Dufwenberg and Patel (2017) additionally investigate the network effects for discrete public goods in this literature.

On the other hand, players with distributive concerns care about relative payoffs, that is, both about own payoff and how it compares with other players' payoffs (their reference level). A significant part of the literature focuses on models in which people exhibit self-centered inequality-aversion (IA): individuals care about their payoff but are also willing to reduce the differences between their payoff and those of the others. If their payoff is lower than the others' (disadvantageous inequality), they envy better-off players and accordingly incur in a disutility. If their payoff is higher than the others' (advantageous inequality), they feel guilt or compassion. Inequality-averse players are assumed to dislike disadvantageous inequality more than advantageous inequality. As a result, they are willing to sacrifice some of their payoffs to reduce either type of inequality, but with a stronger willingness to reduce disadvantageous inequality (Ochs and Roth, 1989; Loewenstein et al., 1989; Bolton, 1991; Kirchsteiger, 1994; Fehr et al., 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Charness and Rabin (2002) introduce an alternative model of distribution concerns in which players maximize social surplus and are also inclined to help other players who receive lower payoffs relative to others.

Lange and Vogt (2003) is the first paper that incorporates IA to IEAs. They study IA in three different games to analyze international environmental cooperation: (i) a symmetric prisoner's dilemma game with binary choice of cooperation or defection; (ii) a symmetric emission game with continuous strategy space; and (iii) a coalition formation game with discrete abatement choice where the second stage is a simultaneous move game (cf. Carraro and Siniscalco, 1993; Hoel, 1992). The authors follow Bolton and Ockenfels' (2000) IA model, with country i's utility defined by:

$$a_i u(y_i) + b_i r(\sigma_i), \tag{11.1}$$

[10] "Paris Deal Would Herald an Important First Step on Climate Change", *The New York Times*, November 29, 2015.

where $a_i, b_i \geq 0$. The first term y_i captures the standard utility from payoff y_i. The second term represents fairness utility based on the relative share $\sigma_i = y_i / \sum_j y_j$. Furthermore, $u(\cdot)$ is differentiable, strictly increasing, and concave; and $r(\cdot)$ is differentiable, concave, and has its maximum at equal share $\sigma_i = 1/N$, where N is the number of countries. The authors find that, depending on its distribution among players, IA can lead to an equilibrium in which majority or all countries cooperate in game (i), since when others cooperate a country values its payoffs to be closer to the equal share, compared to their absolute payoffs. However, in game (ii) with continuous strategies, a country with a possibility to increase its absolute payoff would do so if it receives less than the equal share. Since all countries have such incentives, IA has no effect at all: countries act as if they were maximizing their own payoff in equilibrium. Finally, in game (iii) with coalition formation, countries' strong preference for the equal share leads efficiency gains and even the grand coalition can be stable. In particular, fringe countries outside the coalition incentivize self-interested countries to join and reward the coalition's higher abatement efforts by their own efforts. However, this incentive diminishes as the countries with strong preference for equality join the coalition.

Lange (2006) further analyzes different notions of equity when countries are either industrialized or developing, such as IA with respect to the differences in per capita emissions, and IA with respect to differences in abatement targets. In the model, only industrialized countries take on emission reductions, while developing countries could participate if industrialized countries finance them. He shows that IA concerns on the differences in per capita emissions lead to higher emission reductions in industrialized countries, but no qualitative impact on the incentives to cooperate. On the other hand, IA concerns on the differences between countries' abatement targets and the average of abatement by the other industrialized countries lead (i) developing countries to both participate and abate more, and (ii) higher coalition sizes and even grand coalition if countries are sufficiently IA.

Grüning and Peters (2010) also study the role of fairness in a coalition formation game. In addition to the usual welfare, which consists of benefit minus cost of abating, they subtract the variance in the environmental policy (abatement levels) of all countries. This incentivizes countries to abate similarly and at much higher levels, both within and outside an IEA and, thus, increases the coalition size.

Kolstad (2013) adapts Charness and Rabin's (2002) social preferences to study linear public goods and coalition formation games in which he allows countries to differ in their sizes. More specifically, in his model there are N countries and each country i chooses its abatement level g_i (contribution to public good), or equivalently its emission level $x_i = w_i - g_i$, where w_i is the highest possible emission level. Country i's welfare function has the following form:

$$u_i(x_i, G) = \lambda_i \pi_i(x_i, G) + \delta_i \min_{j \neq i} \pi_j + \varepsilon_i \sum_j \pi_j. \tag{11.2}$$

The first term is the self-interested welfare of country i, with $G = \sum_i g_i$ and a scaling factor λ_i measuring its relative importance for country i. The second term captures the level of country i's care for equity, particularly by the country with the lowest self-interested welfare, scaled by δ_i. Finally, the third term with its scaling factor ε_i captures the preference for efficiency. The author shows that this formulation increases individual

countries' willingness to contribute to the public good. However, it decreases the coalition size. Additionally, heterogeneity in countries' sizes destabilizes coalitions.

Bucholz and Sandler (2016) study the role of other-regarding preferences in public good games with Stackelberg-type leader-follower relations. The standard literature, which is reviewed by the authors, finds that the leaders' unilateral actions to contribute to a public good trigger a reduction in the followers' efforts. However, the authors show that a general other-regarding preference, which captures reciprocity and inequality-aversion, could eliminate this crowding-out effect by leaders' contributions influencing the followers' beliefs positively.

Bucholz and Sandler (2016), and the references cited therein, mention comments by negotiators' highlighting that the IEAs have to be fair and equitable in sharing the burden, as well as emphasizing their willingness to contribute more if others also do so (reciprocity). Lange et al. (2007) and Dannenberg et al. (2010) find empirical support for equity principals and preferences by using data from people involved in international environmental policy. However, Lange et al. (2010) find that equity arguments are often used for self-serving purposes.[11] Therefore, further empirical research is needed to understand the roles of equity concerns and reciprocity on IEAs.

In sum, to the extent that reciprocity and inequity-aversion indeed play a relevant role in IEAs, then they tend to facilitate sustaining an effective agreement and increase the coalition size. However, reciprocity and inequity-aversion can also have detrimental effects, when countries have very high expectations from each other, or when they have different levels of equity-concerns.

7.2 Loss-Aversion and Reference-Dependent Preferences

Kahneman and Tversky's (1979) prospect theory consists of four key elements: i) reference-dependence—the perception of outcomes as gains or losses relative to a reference level; ii) loss-aversion—the tendency towards avoiding losses rather than acquiring gains; iii) diminishing sensitivity—the higher sensitivity to changes around the reference level than to changes away from it; and iv) probability weighting—the subjective overweighting of low probabilities and underweighting of high probabilities. Most of the follow-up literature employs a simplified version of the theory, and utilizes only reference-dependence and loss-aversion (DellaVigna, 2009).

İriş and Tavoni (2018) study the role of loss-aversion and the threat of catastrophic damages, which they call jointly "environmental threshold concerns", on international environmental agreements. They aim to understand whether a threshold for dangerous climate change serves as an effective coordination device for countries to overcome the global free-riding problem. Loss-averse countries decide their abatement level under the threat of either high environmental damages (loss domain), or low damages (gain domain). They find that such concerns can cause the size of the coalition to increase when countries display identical environmental threshold concerns. When countries have heterogeneous threshold concerns, countries with high threshold concerns tend

[11] For more about equity principals and burden sharing rules in international environmental policy, see Cazorla and Toman (2001), Ringius et al. (2002), and Najama et al. (2003).

Table 11.1 Fourfold pattern of risk attitude by Markowitz (1952)

	Small Probability	High Probability
Gains	Risk-seeking	Risk-aversion
Losses	Risk-aversion	Risk-seeking

to join the coalitions, and the coalition size may diminish depending on the level of asymmetry.

İriş (2016) studies loss-aversion and political parties' economic target concerns in an infinitely repeated emission game. Despite optimization's common use in the literature, he argues that political parties often fail to optimize their countries' overall wellbeing since they have additional incentives to be elected. As economic issues influence voters' decisions more than the environmental issues, political parties aim to additionally satisfy their economic targets. He finds that stronger economic target concerns deter the most cooperative emission levels that countries could jointly sustain. Asymmetry either in economic target concerns or in technology levels hinders sustainability. However, when both asymmetries are in effect, they may cancel each other out. The paper concludes that such adverse incentives require efforts at the citizen level to sustain sound international environmental agreements.

Levy (1996) analyzes the impact of prospect theory for international conflict and for bargaining. Following prospect theory that also explains the fourfold pattern of risk attitude (Markowitz, 1952; see Table 11.1), he argues that political parties of adversarial states behave differently when they bargain over losses or gains. Moreover, leaders would be less willing to compromise and more willing to risk large losses to eliminate small losses (diminishing sensitivity).

Similar to Levy (1996), Gsottbauer and van den Bergh (2013) employ prospect theory to investigate how framing by leaders involved in climate agreements affects voters' risk perceptions and preferences for climate policy, in a self-serving way that justifies their standpoint.[12] For instance, the Bush Administration and Al Gore frame climate agreements differently to support their arguments. Specifically, the Bush Administration emphasizes that a strict climate agreement would bring a certain and substantial economic cost, while climate change and its consequences are uncertain. Thus, it frames the climate agreements as a loss, leading to risk-seeking behavior for uncertain losses, and advocates for climate inaction. On the other hand, Al Gore's documentary "An Inconvenient Truth" emphasizes that climate inaction would lead to a high environmental cost and damages, while the economic cost of an agreement is uncertain but should be moderate. Thus, it frames no agreement as a loss, thus justifying support for an effective climate agreement.[13]

Loss-aversion and reference-dependent preferences provide mixed results in terms of their impacts on international environmental cooperation, depending on the framing of the narrative and how the reference levels are determined. Similarly, to almost all the papers employing reference-dependent preferences, the ones reviewed here assume the

[12] Their general objective is to examine the impact of bounded rationality and other-regarding preferences from the perspective of voters who elect the negotiators involved with international climate negotiations.

[13] See the paper for examples of positive frames used by Nordhaus (1992) and Stern (2007).

reference levels to be exogenous. Endogenizing reference levels remains an open question in this literature.

Recently, however, Kőszegi and Rabin (2006) have developed a new model of reference-dependent preferences in which the reference levels are determined endogenously. Specifically, the authors assume that one's reference point is the person's rational expectations about the outcomes, which are determined in a personal equilibrium. The personal equilibrium requires the expectations to be consistent with the optimal behavior, given expectations. Kőszegi and Rabin (2006) and its extensions could provide an interesting agenda for further research on the role of expectations in IEAs.

8. DISCUSSION

We have reviewed both theoretical and experimental papers featuring tipping points and reference dependence, with the aim of extending our understanding of their potentially game-changing impacts on climate change cooperation. To this end, we have examined the role of thresholds and reference levels in public goods and coalition formation games, since they capture important features of dangerous climate change and its impacts on human behavior.

While we have started by highlighting the role of technology in creating strategic complementarities, a key message of this review is that the same holds for the psychological mechanisms described later in the chapter. This is the case because reciprocity leads to strategic complementarity as well (i.e., countries and other actors tend to cooperate conditionally on observable effort by their peers), leading again to a coordination game. The same can be said, to a lesser extent, about inequity aversion, which rationalizes infinitely many equilibria, such that coordination equilibria can be supported (Isaksen et al., 2016). It therefore appears natural to study social and ecological tipping points in a unified framework.

A further reason for integration is that ecological and behavioral tipping points are highly linked. The existence of ecological tipping points associated with abrupt and catastrophic, rather than gradual, climatic change has important behavioral repercussions in terms of the incentives to cooperate on mitigation efforts. Namely, when the threshold for acceptable emission levels, i.e., the amount of emissions that is compatible with gradual change, can be identified with a good degree of confidence, the climate change negotiations can be modeled as a coordination game. Hence, the problem becomes much easier to tackle than in the absence of a tipping point. Intuitively, this is so because the threshold provides an anchor for individuals to coordinate efforts upon. However, uncertainty on the location of the threshold removes the anchor for coordination, and if enough uncertainty surrounds the tipping point the game reverts back to a prisoner's dilemma. This is bad news, since the only self-interested equilibrium in this class of games is defection by all: free-riding incentives lock us into inaction.

An important question is thus which is the best approximation of the real negotiations. One may find reasons for optimism from the fact that negotiators at the COP21 Summit in Paris agreed to take steps towards limiting average global warming to "well below 2°C". Candidate thresholds are thus the symbolic 1.5°C and 2°C targets. However, much uncertainty still plagues the problem of translating such goals into the required actions at the national and subnational level. This is especially important given that the

Paris Agreement lacks legal force and relies instead on pledged nationally determined contributions, which even if fully implemented will not suffice for achieving even the 2°C target. Thus, increased ambition will be needed, in spite of the incentives to delay action.

Non-standard preferences, such as other-regarding and reference-dependent preferences, also have game-changing features that could explain important phenomena regarding climate change. Among the implications that arise from the literature that we have reviewed here are: whether strong leadership in mitigation efforts induces cooperation by others; why countries' high expectations about others' abatement efforts could have detrimental effects; and why developing countries have been relatively reluctant to exert even limited abatement efforts.

Of course, caution must be used when extrapolating from the games reviewed here to real-world issues such as collective action on climate change mitigation. The problem faced by negotiators to international environmental agreements has many more layers of complexity that will make the matter of coordination more difficult. Moreover, implementing agreements, such as the one negotiated in Paris and recently entered into force, that rely on nationally determined pledges, is likely to be further hindered by myopic policymaking. However, introducing realistic behavioral and ecological features, such as reference dependence and tipping points, into mainstream economic modeling appears to be an important step in the right direction.

REFERENCES

Averchenkova, A. and S. Bassi (2016), 'Beyond the targets: assessing the political credibility of pledges for the Paris Agreement', Grantham Research Institute On Climate Change and The Environment, accessed 1 April 2014 at www.lse.ac.uk/GranthamInstitute/publication/beyond-the_targets/.

Barrett, S. (1992), 'International environmental agreements as games', in R. Pethig (ed.), *Conflict and Cooperation in Managing Environmental Resources*, Berlin: Springer-Verlag, pp. 11–37.

Barrett, S. (2005), 'The theory of international environmental agreements', in K.G. Maeler and J. Vincent (eds), *Handbook of Environmental Economics*, Vol. 3, Amsterdam: Elsevier, pp. 1457–516.

Barrett, S. (2013), 'Climate treaties and approaching catastrophes', *Journal of Environmental Economics and Management*, **66**(2), 235–50.

Barrett, S. (2016), 'Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation', *Proceedings of the National Academy of Sciences*, **113**(51), 14515–22.

Barrett, S. and A. Dannenberg (2012), 'Climate negotiations under scientific uncertainty', *Proceedings of the National Academy of Sciences*, **109**(43), 17372–6.

Barrett, S., T.M. Lenton, A. Millner, A. Tavoni, S. Carpenter, J.M. Anderies, C. Folke et al. (2014), 'Climate engineering reconsidered', *Nature Climate Change*, **4**(7), 527.

Bass, F.M. (1969), 'A new product growth for model consumer durables', *Management Science*, **15**(5), 215–27.

Benartzi, S. and R.H. Thaler (1995), 'Myopic loss aversion and the equity premium puzzle', *The Quarterly Journal of Economics*, **110**(1), 73–92.

Biggs, R., S.R. Carpenter and W.A. Brock (2009), 'Turning back from the brink: detecting an impending regime shift in time to avert it', *Proceedings of the National Academy of Sciences*, **106**(3), 826–31.

Bolton, G. (1991), 'A comparative model of bargaining: theory and evidence', *American Economic Review*, **81**, 1096–136.

Bolton, G. and A. Ockenfels (2000), 'A theory of equity, reciprocity, and competition', *American Economic Review*, **90**(1), 166–93.

Bosetti, V., M. Heugues and A. Tavoni (2017), 'Luring others in: coalition formation games with threshold and spillover effects', *Oxford Economics Papers*, **69**(2), 410–31.

Bowman, D., D. Minehart and M. Rabin (1999), 'Loss aversion in a consumption–savings model', *Journal of Economic Behavior & Organization*, **38**(2), 155–78.

Buchholz, W. and T. Sandler (2016), 'Successful leadership in global public good provision: incorporating behavioral approaches', *Environmental and Resource Economics*, **67**(3), 591–607.

Camerer, C.F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press.
Camerer, C.F., G. Loewenstein and M. Rabin (eds) (2003), *Advances in Behavioral Economics*, Princeton, NJ: Princeton University Press.
Carraro, C. and D. Siniscalco (1993), 'Strategies for international protection of the environment', *Journal of Public Economics*, **52**, 309–28.
Cazorla, M.V. and M.A. Toman (2001), 'International equity and climate change policy', in M.A. Toman (ed.), *Climate Change Economics and Policy*, Resources for the Future, Washington, DC, pp. 235–47.
Charness, G. and M. Rabin (2002), 'Understanding social preferences with simple tests', *Quarterly Journal of Economics*, **117**, 817–69.
Ciccarone, G. and E. Marchetti (2013), 'Rational expectations and loss aversion: potential output and welfare implications', *Journal of Economic Behavior & Organization*, **86**, 24–36.
Cox, J.C., D. Friedman and S. Gjerstad (2007), 'A tractable model of reciprocity and fairness', *Games and Economic Behavior*, **59**(1), 17–45.
Currarini, S., C. Marchiori and A. Tavoni (2015), 'Network economics and the environment: insights and perspectives', *Environmental and Resource Economics*, **65**(1), 159–89.
Daido, K., K. Morita, T. Murooka and H. Ogawa (2013), 'Task assignment under agent loss aversion', *Economics Letters*, **121**(1), 35–8.
Dannenberg, A. and A. Tavoni (2016), 'Collective action in dangerous climate change games', in Anabela Botelho (ed.), *WSPC Reference of Natural Resources and Environmental Policy in the Era of Global Change. Vol. 4: Experimental Economics*, World Scientific, pp. 95–120.
Dannenberg, A., B. Sturm and C. Vogt (2010), 'Do equity preferences matter for climate negotiations? An experimental investigation', *Environmental and Resource Economics*, **47**(1), 91–109.
Dannenberg, A., G. Löschel, C. Paolacci, A. Reif and A. Tavoni (2014), 'On the provision of public goods with probabilistic and ambiguous thresholds', *Environmental and Resource Economics*, **61**(3), 365–83.
d'Aspremont, C., A. Jacquemin, J. Gabszewicz and J. Weymark (1983), 'On the stability of collusive price leadership', *Canadian Journal of Economics*, **16**(1), 17–25.
DellaVigna, S. (2009), 'Psychology and economics: evidence from the field', *Journal of Economic Literature*, **47**(2), 315–72.
de Zeeuw, A. (2015), 'International environmental agreements', *Annual Review of Resource Economics*, **7**(1), 151–68.
Diamantoudi, E. and E. Sartzetakis (2006), 'Stable international environmental agreements: an analytical approach', *Journal of Public Economic Theory*, **8**(2), 247–63.
Dixit, A. (2003), 'Clubs with entrapment', *American Economic Review*, **93**(5), 1824–9.
Dufwenberg, M. and G. Kirchsteiger (2004), 'A theory of sequential reciprocity', *Games and Economic Behavior*, **47**(2), 268–98.
Dufwenberg, M. and A. Patel (2017), 'Reciprocity networks and the participation problem', *Games and Economic Behavior*, **101**, 260–72.
Eisenkopf, G. and S. Teyssier (2013), 'Envy and loss aversion in tournaments', *Journal of Economic Psychology*, **34**, 240–55.
Falk, A. and U. Fischbacher (2006), 'A theory of reciprocity', *Games and Economic Behavior*, **54**, 293–315.
Fehr, E. and K. Schmidt (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics*, **114**(3), 817–68.
Fehr, E., G. Kirchsteiger and A. Riedl (1993), 'Does fairness prevent market clearing? An experimental investigation', *Quarterly Journal of Economics*, **108**(2), 437–59.
Fehr, E., S. Gächter and G. Kirchsteiger (1997), 'Reciprocity as a contract enforcement device: experimental evidence', *Econometrica*, **65**, 833–60.
Fehr, E., G. Kirchsteiger and A. Riedl (1998), 'Gift exchange and reciprocity in competitive experimental markets', *European Economic Review*, **42**, 1–34.
Finus, M. (2008), 'Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges', *International Review of Environmental and Resource Economics*, **2**(1), 29–67.
Freund, C. and Ç. Özden (2008), 'Trade policy and loss aversion', *The American Economic Review*, **98**(4), 1675–91.
Genesove, D. and C. Mayer (2001), 'Loss aversion and seller behavior: evidence from the housing market', *The Quarterly Journal of Economics*, **116**(4), 1233–60.
Gladwell, M. (2000), *The Tipping Point: How Little Things Make a Big Difference*, Boston, MA: Little, Brown.
Granovetter, M. (1978), 'Threshold models of collective behavior', *The American Journal of Sociology*, **83**(6), 1420–43.
Greene, D.L. (2011), 'Uncertainty, loss aversion, and markets for energy efficiency', *Energy Economics*, **33**(4), 608–16.

Grüning, C. and W. Peters (2010), 'Can justice and fairness enlarge the size of international environmental agreements?', *Games*, **1**, 137–58.
Gsottbauer, E. and J.C. van den Bergh (2013), 'Bounded rationality and social interaction in negotiating a climate agreement', *International Environmental Agreements: Politics, Law and Economics*, **13**(3), 225–49.
Hadjiyiannis, C., D. İriş and C. Tabakis (2012), 'International environmental cooperation under fairness and reciprocity', *The B.E. Journal of Economic Analysis & Policy*, **12**(1), 1–30.
Heal, G. and H. Kunreuther (2012), 'Managing catastrophic risk', Paper No. w18136, National Bureau of Economic Research.
Heath, C., R. Larrick and G. Wu (1999), 'Goals as reference points', *Cognitive Psychology*, **38**(1), 79–109.
Hoel, M. (1992), 'International environment conventions: the case of uniform reductions of emissions', *Environmental and Resource Economics*, **2**(2), 141–59.
IPCC (2013), 'Summary for policymakers, in Climate Change 2013: The Physical Science Basis', in T.F. Stocker et al. (eds), *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge: Cambridge University Press.
İriş, D. (2016), 'Economic targets and loss-aversion in international environmental cooperation', *Journal of Economic Surveys*, **30**(3), 624–48.
İriş, D. and A. Tavoni (2018), 'Loss aversion in international environmental agreements', *Environmental and Resource Economics Review*, **27**(2), 363–97.
İriş, D., J. Lee and A. Tavoni (2019), 'Delegation and public pressure in a threshold public goods game', *Environmental and Resource Economics*, **74**(3), 1331–53.
Isaksen, E., A. Richter and K.A. Brekke (2016), 'When kindness generates unkindness. Why positive framing cannot solve the tragedy of the commons', presented at EAERE 2016.
Jang, D., A. Patel and M. Dufwenberg (2016), 'Reciprocity and agreements', unpublished manuscript.
Kagel, J.H. and A.E. Roth (eds) (1995), *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press.
Kahneman, D. (2003), 'A psychological perspective on economics', *The American Economic Review*, **93**(2), 162–8.
Kahneman, D. and A. Tversky (1979), 'Prospect theory: an analysis of decision under risk', *Econometrica*, **47**(2), 263–91.
Kahneman, D. and A. Tversky (eds) (2000), *Choices, Values, and Frames*, New York: Cambridge University Press.
Kahneman, D., J.L. Knetsch and R.H. Thaler (1986), 'Fairness as a constraint on profit seeking: entitlements in the market', *American Economic Review*, **76**(4), 728–41.
Kahneman, D., J.L. Knetsch and R.H. Thaler (1991), 'Anomalies: the endowment effect, loss aversion, and status quo bias', *Journal of Economic Perspectives*, **5**(1), 193–206.
Kirchsteiger, G. (1994), 'The role of envy in ultimatum games', *Journal of Economic Behavior and Organization*, **25**, 373–89.
Köbberling, V. and P.P. Wakker (2005), 'An index of loss aversion', *Journal of Economic Theory*, **122**(1), 119–31.
Kolstad, C.K. (2013), 'International environmental agreements with other-regarding preferences', unpublished paper, Stanford University.
Kőszegi, B. and M. Rabin (2006), 'A model of reference-dependent preferences', *The Quarterly Journal of Economics*, **121**(4), 1133–65.
Lade, S., A. Tavoni, S. Levin and M. Schlüter (2013), 'Regime shifts in a social-ecological system', *Theoretical Ecology*, **6**, 359–72.
Lange, A. (2006), 'The impact of equity-preferences on the stability of international environmental agreements', *Environmental and Resource Economics*, **34**, 247–67.
Lange, A. and C. Vogt (2003), 'Cooperation in international environmental negotiations due to a preference for equity', *Journal of Public Economics*, **87**, 2049–67.
Lange, A., C. Vogt and A. Ziegler (2007), 'On the importance of equity in international climate policy: an empirical analysis', *Energy Economics*, **29**, 545–62.
Lange, A., A. Löschel, C. Vogt and A. Ziegler (2010), 'On the self-serving use of equity principles in international climate negotiations', *European Economics Review*, **54**, 359–75.
Leibenstein, H. (1950), 'Bandwagon, snob, and Veblen effects in the theory of consumers' demand', *Quarterly Journal of Economics*, **64**(2), 183–207.
Lenton, T.M. (2011), 'Early warning of climate tipping points', *Nature Climate Change*, **1**(4), 201–9.
Lenton, T., H. Held, E. Kriegler, J.W. Hall, W. Lucht, S. Rahmstorf and H.J. Schellnhuber (2008), 'Tipping elements in the Earth's climate system', *Proceedings of the National Academy of Sciences*, **105**(6), 1786–93.
Levy, J. (1996), 'Loss aversion, framing, and bargaining: the implications of prospect theory for international conflict', *International Political Science Review*, **17**(2), 179–95.
Loewenstein, G., L. Thompson and M. Bazerman (1989), 'Social utility and decision making in interpersonal contexts', *Journal of Personality and Social Psychology*, **57**, 426–41.

Marchiori, C., S. Dietz and A, Tavoni (2017), 'Domestic politics and the formation of international environmental agreements', *Journal of Environmental Economics and Management*, **81**, 115–31.
Markowitz, H. (1952), 'The utility of wealth', *Journal of Political Economy*, **60**, 151–8.
Milinski, M., R.D. Sommerfeld, H.J. Krambeck, F.A. Reed and J. Marotzke (2008), 'The collective-risk social dilemma and the prevention of simulated dangerous climate change', *Proceedings of the National Academy of Sciences*, **105**(7), 2291–4.
Milinski, M., C. Hilbe, D. Semmann, R. Sommerfeld and J. Marotzke (2016), 'Humans choose representatives who enforce cooperation in social dilemmas through extortion', *Nature Communications*, **7**, 10915.
Najama, A., S. Huq and Y. Sokona (2003), 'Climate negotiations beyond Kyoto: developing countries concerns and interests', *Climate Policy*, **3**, 221–31.
Nordhaus, W.D. (1992), 'The "DICE" model: background and structure of a dynamic integrated climate economy model of the economics of global warming', Cowles Foundation Discussion Paper No. 1009.
Nordhaus, W.D. (2008), *A Question of Balance: Weighing the Options on Global Warming Policies*, New Haven, CT and London: Yale University Press.
Nordhaus, W.D. (2015), 'Climate clubs: overcoming free-riding in international climate policy', *American Economic Review*, **105**(4), 1339–70.
Nyborg, K. (2015), 'Reciprocal climate negotiators', IZA Discussion Paper No. 8866.
Ochea M. and A. de Zeeuw (2015), 'Evolution of reciprocity in asymmetric international environmental negotiations', *Environmental and Resource Economics*, **62**(4), 837–54.
Ochs, J. and A.E. Roth (1989), 'An experimental study of sequential bargaining', *American Economic Review*, **79**, 355–84.
Ostrom, E. (2009), 'A polycentric approach for coping with climate change', World Bank Policy Research Working Paper 5095, Washington, DC: World Bank.
Rabin, M. (1993), 'Incorporating fairness into game theory and economics', *American Economic Review*, **83**(5), 1281–302.
Ringius, L., A. Torvanger and A. Underdal (2002), 'Burden sharing and fairness principles in international climate policy', *International Environmental Agreements: Politics, Law and Economics*, **2**, 1–22.
Rockström, J., W. Steffen, K. Noone, A. Persson, F.S. Chapin III, E.F. Lambin and B. Nykvist et al. (2009), 'A safe operating space for humanity', *Nature*, **461**, 472–5.
Rogers, E.M. (2003), *Diffusion of Innovations*, New York: Free Press.
Scheffer, M. and S.R. Carpenter (2003), 'Catastrophic regime shifts in ecosystems: linking theory to observation', *Trends in Ecology & Evolution*, **18**, 648–56.
Scheffer, M., S. Carpenter, J.A. Foley, C. Folke and B. Walker (2001), 'Catastrophic shifts in ecosystems', *Nature*, **413**, 591–6.
Segal, U. and J. Sobel (2007), 'Tit for tat: foundations of preferences for reciprocity in strategic settings', *Journal of Economic Theory*, **136**, 197–216.
Shafir, E., P. Diamond and A. Tversky (1997), 'Money illusion', *The Quarterly Journal of Economics*, **112**(2), 341–74.
Stern, N. (2007), *The Economics of Climate Change: The Stern Review*, Cambridge: Cambridge University Press.
Tavoni, A. and S. Levin (2014), 'Managing the climate commons at the nexus of ecology, behaviour and economics', *Nature Climate Change*, **4**, 1057–63.
Tavoni, A., A. Dannenberg, G. Kallis and A. Löschel (2011), 'Inequality, communication and the avoidance of disastrous climate change in a public goods game', *Proceedings of the National Academy of Sciences*, **108**(29), 11825–9.
Tovar, P. (2009), 'The effects of loss aversion on trade policy: theory and evidence', *Journal of International Economics*, **78**(1), 154–67.
Tversky, A. and D. Kahneman (1991), 'Loss aversion in riskless choice: a reference-dependent model', *The Quarterly Journal of Economics*, **106**(4), 1039–61.
Vasconcelos, V.V., F.C. Santos, J.M. Pacheco and S.A. Levin (2014), 'Climate policies under wealth inequality', *Proceedings of the National Academy of Sciences*, **111**(6), 2212–16.
Young, H.P. (2009), 'Innovation diffusion in heterogeneous populations: contagion, social influence and social learning', *American Economic Review*, **99**, 1899–924.
Watts, D.J. (2002), 'A simple model of global cascades on random networks', *Proceedings of the National Academy of Sciences*, **99**(9), 5766–71.
Weibull, J.W. (1997), *Evolutionary Game Theory*, Cambridge, MA: MIT Press.
Weikard, H., L. Wangler and A. Freytag (2014), 'Minimum participation rules with heterogeneous countries', *Environmental and Resource Economics*, **62**(4), 711–27.
Weitzman, Martin L. (2009), 'On modeling and interpreting the economics of catastrophic climate change', *Review of Economics and Statistics*, **91**(1), 1–19.