Coordination games vs prisoner's dilemma in sustainability games: a critique of recent contributions and a discussion of policy implications

Alessio Carrozzo Magli[1] and Piero Manfredi[2]

[1]Alessio Carrozzo-Magli, Dipartimento di Economia, Università di Bologna, Piazza Scaravilli 2, 40126 Bologna, ITALY, e-mail: alessio.carrozzo2@unibo.it (corresponding author).

[2]Piero Manfredi, Dipartimento di Economia & Management, Università di Pisa, Via Ridolfi 10, 56124 Pisa (Italy).

Abstract

Recent works have suggested that most games arising in climate change diplomacy and sustainability choices, should have a coordinative nature rather than that of a prisoners' dilemma, as typically suggested. In this note, after having proposed a definition of sustainability game, we critically review the merits and shortcomings of these contributions and use a simple, yet sufficiently general, model to recall the difficulties for coordination to emerge in such games. Indeed, as far as the players' short-term interest is involved, at least in some degree, these games will most often generate a prisoner's dilemma, thereby allowing coordination only upon long-term interactions possibly under the pressure of a continuing environmental deterioration. A counter-intuitive result is proved, showing the circumstances when the deterioration of the environment can hinder cooperation in repeated games. We conclude by highlighting a number of factors forcing "brown" behaviour and therefore threatening coordination, first of all poverty and inequality, and pinpointing that, though ability to enact coordination will be key for a successful battle for climate, undue emphasis on coordination might be the deleterious in view of its optimistic message.

Keywords. Environmental degradation, climate change, sustainability game, prisoner's dilemma, coordination.

1. Introduction

Recent contributions in this Journal (Decanio and Fremstand, 2013, since now on DF2013 for brevity; Mielke and Steudle, 2018, MS2018 for brevity) have argued that the strategic interactions underlying climate negotiations and in general most sustainability games, would primarily have a coordination nature, in contrast with a more traditional view describing them as a prisoner's dilemma, under a variety of approaches (Carraro and Siniscalco 1993; Wang et al 2009; Heitzig et al 2011; Hugues 2013; Nordhaus 2015; Mielke and Steudle, 2018 and refs therein).

In particular, FD2013 compared the different 2x2 one-shot games relevant for international bilateral climate negotiations and eventually concluded that a world of "reasonable" people, trusting the authority of science on the seriousness of the climate threat, would lie on a coordination game. On a different line resting on a time-horizon argument, namely that the serious effects of global climate change are expected in a long-term future while the time horizon of actual investors having to choose right now between brown and green technologies is much shorter, MS2018 also claimed, still by a static framework, that investments in mitigation and adaptation will likely take the form of a coordination problem. One of their main argument is that such investments will be more profitable the more the other players are investing in green technologies.

By this note, we aim at critically assessing these novel insights on the coordinative nature of sustainability games with respect to the traditional view that mostly classifies them as a prisoner's dilemma (Nordhaus 2015; Hugues 2013; Heitzig et al 2011; Wang et al 2009; Carraro and Siniscalco 1993). Our point is that the results in DF2013 and MS2018 though correct per se, are limited by the specific foci of their hypotheses. Reconciliation readily emerges, in a broader perspective, as soon as one attempts at including more general hypotheses and some long-term character in their frameworks. Consistently, we develop this note by the following steps. After having supplied a definition of "sustainability game", we critically review the cited works and highlight their possible shortcomings. Next, we consider a simple game-theoretic setup yet broadly encompassing the cited works but under more general hypotheses, in order to adequately inform the discussion on the dispute "prisoner's dilemma vs coordination" and compare its short-term outcomes with its longer-term implications in relation to sustainability games. In this setup, we also include the progressive deterioration of the environment (IPCC, 2021) over time as a necessary characteristic of sustainability games.

Our main claim is that, unlike the cited contributions, the nature of prisoner's dilemma of most sustainability games cannot be easily removed. Indeed, even if global climate change might eventually be universally acknowledged as a critical threat to human activity, this will hardly cancel the short run individual incentive to be "brown" as far as some individual self-interest persists in the rules of the game. Clearly, if the real (and perceived) threat should blow-up, defecting (and polluting) would become too risky, incentivizing cooperation. But this incentive can only emerge in a repeated game, where the history of past interactions matters in determining current and future behaviour. However, the effects of the continued environment deterioration - as forecasted by climate science (IPCC 2021) - are shown not to be univocal as, under certain conditions, they can hinder rather than favour coordination.

This note is organised as follows. In section 2, we critically review the cited works by DF2013 and MS2018. In section 3, we provide our discussion by first introducing a basic 2-agents setup, which is then extended to a multi-agent case. Concluding remarks follow in Section 4.

2. Sustainability games and coordination: a definition and a critique of some recent efforts

Though strategic interactions affecting environment and climate (termed "mitigation games" in Heitzig et al., 2011), and in general the "sustainability" issue, are pervasive in the cited literature, they do not seem to have been precisely defined. By a *sustainability game* we intuitively mean any strategic interactions - involving any type of economic, social and political agents (consumers and producers, intermediate societal bodies and institutions, governments) - specifically dealing with environment preservation *latu sensu*, and whose outcome could relevantly impact on the Earth's fundamental resources available to current and future generations. More technically, a sustainability game can be considered as a public- good game with a number of additional characteristics resulting from the massive evidence on the climate emergency (IPCC 2021) namely: (i) *time irreversibility*, that is every new shot of the game will be played under worsened conditions, due to the continued degradation of the environment. Even if certain agents/sectors do not suffer such effects in the short-term, they will do shortly after e.g., in the form of stricter regulations; (ii) *pervasiveness*, i.e., environmental degradation will always affect all players' payoffs, (iii) *globality*, i.e., every defection in a given shot, and ensuing environmental damage, will always extend to a broader scale than the "local" one involving the agents actually competing (iv) *"non-resilience to unaffordability"*, i.e., the presence of significant heterogeneities in resource endowments will always undermine cooperation. Borrowing MS2018 argument of the asymmetry of horizons, it might happen that especially in the current epoch (i),(ii),(iii) hold only weakly for specific groups. In this case the incentive to defect would persist and the prisoner's dilemma nature of the game would prevail even more so.

Here, motivated by DF2013 and MS2018, we comment about simple game-theoretic models of two specific, yet broad, types of sustainability games, one at a top level (countries' bilateral negotiations), the other one at a micro-level (firms' investments in green technologies).

2.1 Coordination in climate change diplomacy

Focusing on climate change diplomacy, DF2013 nicely analysed all the (many dozens) of 2x2 static, one-shot, games, and identified those potentially suited (25) to represent bilateral climate negotiations between countries, as well as the conditions making them fit the actual state of the world. A particular effort was devoted to the dispute "prisoner's dilemma vs coordination".

Their main claim is that such games cannot but have a coordinative nature provided that scientific knowledge is perfectly spread among players and there is a non-zero probability that catastrophic threats are triggered by economic activities. They also pinpointed that, if not even such an epochal threat is perceived as more important than own selfish advantage, the world would never escape from the prisoner's dilemma logic.

Though the previous statements are self-evident, at least three game-based considerations are necessary in view of their potential policy implications.

First, were climate negotiations a pure matter of coordination, only communication failures would prevent cooperation (i.e., the fear that the others will not cooperate). Empirically, failures of international climate agreements would not be expected to be the rule because every joining country would credibly communicate its intention to cooperate and there would be no reason to fear defections, as they would be individually unprofitable.
Second, even in DF2013 "2-big players" ("powers") setting, the non-zero risk of a climate catastrophe has no room in a one-shot game, whereas it becomes more and more relevant as one extends the time horizon. One shot static game theory is only a matter of finding the best response to the other players' strategies. If defecting and "pollute" when all the others "abate" is profitable (and if, more trivially, polluting is better than abating when all the others are polluting), then all players will pollute. Therefore, if players face an individual incentive to defect, that is, the one-shot game is a prisoner's dilemma, cooperation can result as a stable equilibrium only in presence of many (infinite or indefinite) repeated interactions (this is just the textbook Folk theorem). Indeed, in this case there would be enough time to punish defectors (i.e., polluters) at every stage of the game and therefore defecting would prevent players from benefiting an infinite stream of cooperative (Pareto Superior) payoffs.
Last, the current global geopolitical chessboard suffers a much greater complexity, in terms of both the number of players and their relative "power", than the (symmetric) 2-big –player setting considered in DF2013. The lower the individual impact, the less likely the individual player will be pivotal in determining a climate catastrophe (and more in general, the lower the individual impact the higher the incentive to defect). This difficulty in coordination is well-known from basic theory and has been widely demonstrated since Rapoport's seminal experiments on the role of an increasing number of agents in hindering coordination (Rapoport 1988, Cadsby and Maynes 1999).

2.2 Green investments, coordination and determinants of coordination failure
MS 2018 suggested that most sustainability games proposed in the literature have been represented as either a Tragedy of the commons (Diekert 2013; Pachauri 2014) or as a prisoner's dilemma[1] (Carraro and Siniscalco 1993; Wang et al 2009; Heitzig et al 2011; Hugues 2013; Nordhaus, 2015; MS 2018 and refs therein) because they focused on a time-asymmetric comparison, namely between short-term mitigation costs and the long-term benefits of the avoidance of the damages of climate change. However, many sustainability games e.g., those related to green investment "are rarely motivated by physical climate damages in the future, but by achieving returns in the present… Also, adding to technical progress and spillover effects, the promise of global benefits can influence investment decisions… can take the form of a stag hunt." (MS 2018)
We introduce here a few details about MS2018 formulation as useful for our subsequent developments because, though focusing on a specific topic (green vs brown investments), it is general enough to describe most types of sustainability games. They considered a community of n symmetric players ("investors") such that: (i) if player i invests a given amount of money I_i in

[1] Since it was proved that the standard Hardin's formulation of the Tragedy of the Commons (Hardin, 1968) represents a proper prisoner's dilemma (Carrozzo Magli et al, 2021), we will not keep this distinction and only speak of prisoners' dilemmas.

green technology, her payoff is $gI_i + \frac{\gamma}{n}\sum_{j=1}^{m} I_j$, where $g = r_g - c_g$ is the (constant) net return resulting from the individual investment, and m is the number of players investing green. The second term reflects the social legitimation of the green technology as enforced by its diffusion, by which a higher (relative) number (m) of players investing "green" makes this technology collectively more profitable, according to a factor γ; (ii) If player i invests the same amount in brown technology, her payoff is $bI_i + \frac{\beta}{n}\sum_{j=1}^{n-m} I_j$. Consistently with their aims, MS2018 postulate $b > g > 0$ (i.e., the brown technology is more profitable in the absence of diffusion of the green one) and $\gamma > \beta > 0$ (the green technology becomes profitable when widespread). Notably, this game is not necessarily a coordination (stag hunt) game. For this to occur, another assumption (not explicitly stated in MS2018) is necessary, namely that one shot defection to the brown investment when all players are playing green is not profitable i.e.,

$$bI + \frac{\beta}{n}I < gI + \frac{\gamma}{n}nI \rightarrow b + \frac{\beta}{n} < g + \gamma \qquad (1)$$

Notably, were condition (1) not satisfied, players would play a "trivial" game. The implication of (1) is far from irrelevant: if the game is a stag hunt, then playing brown is dominated by the green strategy only when the majority of players is green, but brown players' payoff starts increasing as more agents are playing brown. This corresponds to postulating that in an economy populated by brown agents only, the difference between the spillovers generated in the investment market and the overall environmental damage linearly increases in the number of brown players, bringing to the striking conclusion that *the higher the number of polluters, the better for everyone and the lower the impact of the environmental damage inherently characterizing brown technologies*. The same holds if the game is trivial: in this case, being all brown is even better than being all green, a conclusion even stronger than those implicit in the prisoner's dilemma.

However, the key feature of MS2018 game is that, by postulating $\beta \geq 0$, they implicitly prevent *a priori* the game from being a prisoner's dilemma. Said otherwise, their game is not a public good game (and consequently not a sustainability one). Rather it represents a strategic interaction between adopters of two technologies where the greener one, or the less brown, has the potential to spread (from which their relevant conclusion that the main cause of coordination failure would be the green technology riskiness).

3. Prisoner's dilemma vs cooperation in sustainability games

3.1 A basic 2x2 setup

We setup our discussion by setting the MS2018 framework and notations within a generic sustainability game. This setup can be readily adapted, *mutatis mutandis*, to study most types of strategic interactions relevant for sustainability games. We start from a baseline 2-player model that will later extended to a multi-player case. As a first step, we restore in the simplest manner the hypothesis that the considered games are standard public good games, by assuming that the diffusion of the brown technology imposes cost distributed on the entire community including green players (i.e., the aforementioned pervasiveness hypothesis), by taking $\beta < 0$ (which is not in

contrast with MS2018 hypothesis that $\gamma > \beta$)[2]. Additionally, we maintain $b > g$ and keep a positive γ factor to stick to the original MS2018 formulation, though this parameter is irrelevant for subsequent developments. A similar linear damage function was empirically tested by Dell, Jones and Olsen (2008) and adopted e.g., by Pindyck (2012).

The one-shot game

For notational simplicity, in the resulting payoff matrix (Table 1) we subtracted a quantity β>0 rather than adding a negative one. Accordingly, if $b - g - \gamma > \frac{\beta}{2}$ i.e., if the private benefit from playing brown is greater than the social cost this behaviour imposes to the community, the resulting game is a prisoner's dilemma with (brown, brown) as unique Nash equilibrium (we disregard the extreme case $b - g - \gamma > \beta$ where a full brown world is best for all agents). Only when $b - g - \gamma < \frac{\beta}{2}$ the coordinative equilibrium (green, green) emerges and the game is a stag hunt (or, in the limit case in which $b - g - \frac{\gamma}{2} < \frac{\beta}{2}$, a trivial game in which being green is the best response even if all the other players are playing brown).

	Player 2	
Player 1	Green	Brown
Green	$g + \gamma, g + \gamma$	$g + \frac{\gamma - \beta}{2}, b - \frac{\beta}{2}$
Brown	$b - \frac{\beta}{2}, g + \frac{\gamma - \beta}{2}$	$b - \beta, b - \beta$

Table 1. Payoff matrix in the basic strategic interaction between two producers.

The repeated game: standard facts

As regards the issue of market failure (arising in the prisoner's dilemma case: $b - g - \gamma > \frac{\beta}{2}$), textbook-level game theory provides a number of solutions, ranging from (Pigouvian) taxes on defectors in the one-shot game up to the use of the grim trigger strategy along repeated games. For example, letting δ $(0 < \delta < 1)$ denote the inter-shot discount factor, the Folk theorem states that agents will cooperate by choosing the green process, for sufficiently large discount rates i.e., for:

$$\frac{g+\gamma}{(1-\delta)} > b - \frac{\beta}{2} + \frac{\delta}{1-\delta}(b - \beta) \quad \rightarrow \quad \delta > \frac{b-g-\gamma-\frac{\beta}{2}}{\frac{\beta}{2}}. \qquad (2)$$

Briefly, cooperation can emerge (note that the numerator of (2) is always positive in the prisoner's dilemma region) only if players assign a sufficiently high value to future profit (i.e., provided that they are not myopic towards the future), which will be possible only if the environment is preserved.

[2] A realistic formulation should include both (i) a positive "legitimation" effect of playing brown (i.e., if the majority plays brown this will increase the payoff of all those playing brown) and (ii) a social cost of playing brown, possibly inverted-U shaped mirroring the fact that at high levels of diffusion of the brown technology the ensuing environmental damage will overtake, at the individual level, the benefit of social legitimation. As this richness is not necessary for our argument, we keep the formulation as parsimonious as possible.

Counter-intuitive implications of nature degradation and repeated games

Let us now include the *time-irreversibility* property, owing to the large available empirical evidence that any defection (and/or delay of implementation of the appropriate policy interventions) will lead to further deteriorated environmental conditions (IPCC 2021, Meadows et al. 2004). This will possibly act directly, by expanding the direct environmental damage (β), and also indirectly by e.g., reducing availability of raw materials, increasing costs, and eventually reducing profitability. Here, we assume for simplicity that this "Nature discount rate"[3] has the form $\delta_N = e^{-r\tau}$, where $\tau > 0$ is the time distance between consecutive shots and r the rate of environment degradation per unit time, and it negatively affects b, γ and g. For parsimony we do not include a positive effect on β (our results would hold *a fortiori* in this case) and take (without loss of generality) $\tau = 1$.

If agents continue to play, in each time shot t, the only Nash equilibrium (brown,brown), then the environment degradation will continue up to the point where individual defection will not be convenient anymore i.e., $e^{-rt}(b - g - \gamma) < \frac{\beta}{2}$, and the game loses its nature of prisoner's dilemma, evolving either into (i) a coordination game if $e^{-rt}\left(b - g - \frac{\gamma}{2}\right) > \frac{\beta}{2}$, or (iii) a trivial game otherwise. In the coordination case (i), basic theory tells that the two (pure strategy) Nash equilibria (green, green) and (brown, brown) are both locally stable, and there is an unstable mixed strategy equilibrium where each player chooses to play green with probability $q = \frac{1}{\gamma}\left(2\left(b - g - \frac{\gamma}{2}\right) - \beta e^{rt}\right)$.

If players are not able to escape from the coordination failure, b, g and γ (and therefore q) will continue to decrease up to the point where the game degenerates into a trivial game, making the green outcome to eventually emerge. Briefly, persistent environmental degradation will eventually force coordination, but this might occur after many shots yielding to a (much) more degraded environment that is, payoffs would be higher had players been green from the beginning.

	Player 2	
Player 1	Green	Brown
Green	$e^{-rt}(g+\gamma), e^{-rt}(g+\gamma)$	$e^{-rt}\left(g+\frac{\gamma}{2}\right)-\frac{\beta}{2}, e^{-rt}b-\frac{\beta}{2}$
Brown	$e^{-rt}b-\frac{\beta}{2}, e^{-rt}\left(g+\frac{\gamma}{2}\right)-\frac{\beta}{2}$	$e^{-rt}b-\beta, e^{-rt}b-\beta$

Table 2: Payoff matrix in the basic strategic interaction under continued environment degradation.

A subtle consequence of this state of affairs emerges when one leaves the previous simplistic setting of subsequent disconnected one-shot games, and correctly considers a repeated game setting. In this case, it happens that threatening a grim trigger strategy is no longer credible. The reason is that when (brown,brown) ceases to represent a Nash equilibrium, players cannot anymore credibly commit to play brown because it has become a sub-optimal strategy. Consider the simplest possible situation where the game evolves into case (ii) after one single shot of

[3] The issues of discounting, uncertainty and myopia in environmental analyses are far-reaching ones (see Heal 2007 and refs therein; Polasky et al, 2019 and refs therein). Here, we just kept the approach as simple as possible.

environment degradation, that is, $e^{-r}\left(b-g-\frac{\gamma}{2}\right)<\frac{\beta}{2}$. The only credible threat in the repeated game is therefore playing brown after observing a defection, and then playing (in the game represented in Table 2) green forever. The correct condition for cooperation on the green outcome at the initial shot (i.e., before any environmental degradation) takes the form:

$$\frac{g+\gamma}{(1-\delta)}>b-\frac{\beta}{2}+(b-\beta)\delta+\frac{e^{-r}\delta^2}{1-\delta}(g+\gamma) \qquad (3)$$

Comparing (3) with the analogous condition (2) in the absence of environment degradation, one notes – given that the left-hand sides are identical – that if the right member of (3) is greater than the corresponding term in (2), the environment degradation will frustrate cooperation rather than favouring it, making more likely that the final outcome of the game becomes an infinite sequence of deteriorated games as represented in Table 2. Formally this occurs under the simple condition

$$e^{-r}(g+\gamma)>b-\beta \quad (4)$$

which is fulfilled for not too large magnitudes of the Nature discount rate.
Briefly, environment degradation scales down payoffs but, at the same time, it prevents players from inflicting each other an infinite sequence of punishments. These results suggest that the implications of including the nature discount rate are far from trivial.

3.2 A multi-agent framework

We now extend previous ideas within MS2018 multi-agent formulation to argue that the issues of the 2x2 framework persist in the multi-agent setting, possibly in an aggravated form. We relax the linearity of payoffs in the amount invested, by taking (standard) quadratic cost functions in the form: $c_B=\frac{c}{2}B_i^2$, $c_G=\frac{d}{2}G_i^2$, where B_i (G_i) is the amount invested in the brown (green) technology. We keep the assumption that playing brown is cheaper: $c<d$[4].

The one-shot game

Brown investors decide the amount of money B_i to be invested in order to maximize the payoff function:

$$S_b=B_i-\frac{c}{2}B_i^2-\frac{\beta B}{n} \ , \ B=\sum_i^n B_i$$

The maximization problem yields:

$$B_i^*=\frac{n-\beta}{cn} \quad \rightarrow \quad S_b^*=\frac{(n-\beta)^2}{2cn^2}-\frac{\beta}{n}B_{-i},$$

where $B_{-i}=B-\sum_{j\neq i}^n B_j$ represents brown investment from all players but i.
Exploiting symmetry, if all players play brown each one gets a payoff:

$$S_b^*=\frac{(n-\beta)[n-\beta(2n-1)]}{2cn^2}\xrightarrow[n\to\infty]{}\frac{1-2\beta}{2c}$$

[4] In a standard Bertrand Oligopoly model, this assumption is sufficient for brown producers to completely wipe out green competitor from the market, by simply charging a price lower than the green marginal cost.

Defecting when all are playing green yields the following pair:

$$B_d = \frac{n-\beta}{nc} \quad , \quad S_d = \frac{(n-\beta)^2}{2cn^2} \xrightarrow[n\to\infty]{} \frac{1}{2c}$$

Notably, in contrast to Mielke and Steudle, a full brown configuration yields lower payoff than those obtained by a single brown deviator i.e., $S_d > S_b^*$.

Green investors choose the amount of money to maximize the payoff function:

$$S_g = G_i + \gamma\frac{G}{n} - \frac{d}{2}G_i^2 - \frac{\beta B}{n}$$

Maximizing this objective function and exploiting symmetry, if all agents play green, one gets:

$$G_i^* = \frac{n+\gamma}{dn}$$

and

$$S_g^* = \frac{(n+\gamma)^2}{2dn^2} + \frac{\gamma}{n}G_{-i} = \frac{(n+\gamma)[n+(2n-1)\gamma]}{2dn^2} \xrightarrow[n\to\infty]{} \frac{2\gamma+1}{2d}$$

In order the present game be a coordination game, individual defection must be unprofitable, yielding the following pair of conditions

$$\frac{(n+\gamma)[n+(2n-1)\gamma]}{2dn^2} > \frac{(n-\beta)^2}{2cn^2}$$

$$\frac{d}{c} < \frac{(n+\gamma)[n+(2n-1)\gamma]}{(n-\beta)^2} \xrightarrow[n\to\infty]{} 2\gamma+1$$

This means that only when the green technology is not particularly expensive with respect to the brown one, the resulting game is a coordination rather than a prisoner's dilemma. But brown technologies are adopted exactly because they are cheaper, suggesting that the above inequality may not be satisfied[5].

Figure 1 provides a graphic summary of our main results for the case of a large population of agents ($n \to \infty$). The outcome of the game can be represented compactly in terms of the relative cost $T = \frac{d}{c}$ of the green technology by a few thresholds $T_G = 1 < T_C < T_P$, expressed in terms of the other model parameters (further details in the Appendix). In particular, the outcome will be (i) dominance of the brown technology for very large relative costs of the green one ($T > T_P$), (ii) prisoner dilemma in a first intermediate window of T values ($T_C < T < T_P$), switch to (iii) bistability and coordination by further decreasing T ($T_G < T < T_C$), up to eventually ending into (iv) full dominance of the green technology when the latter becomes even more economically convenient of the brown one ($T < T_G = 1$). Note that threshold T_P tends to T_C when $\beta \to 0$ i.e., when no environmental damage results from the brown technology, and that T_C tends to T_G when $\gamma \to 0$.

[5] Clearly, a very brown technology (i.e., a high β) makes the emergence of coordination easier, but brown technologies are unsustainable when adopted by many agents: the individual impact on the environment is close to 0.

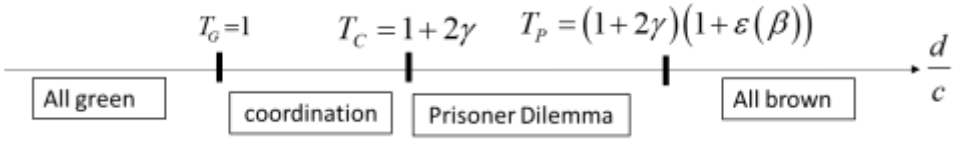

Figure 1. Outcomes of the multi-agent game depending on the values of the relative cost d/c of the green technology.

As a final remark, consider a general production processes involving many intermediate steps and inputs. If there are x intermediate inputs, for each $h = 1, \dots, x$ a generalised condition in the form $T_h = \frac{d_h}{c_h} < 2\gamma_h + 1 = T_{C,h}$ will hold. Notably, if for any of those x markets the brown alternative is sufficiently cheaper, buying a brown input would be the best option even when, on the top level, a green production process is adopted. In this case, the whole game would be a prisoner's dilemma, simply with some "green washing".

The repeated game

With respect to the previous game, let us focus on the case $T_C < T < T_P$, where in particular $\frac{d}{c} > 2\gamma + 1$, implying the one-shot game is a standard prisoner's dilemma. Then, agents will cooperate by choosing the green process, for sufficiently large discount rates i.e., for:

$$\frac{2\gamma+1}{2d(1-\delta)} > \frac{1}{2c} + \frac{\delta}{1-\delta}\frac{1-2\beta}{2c} \quad \rightarrow \quad \delta > \frac{d}{c(2\gamma+1)+2d\beta}$$

Including also the Nature discount rate, the condition for cooperation would depend on the number of shots necessary for the game to degenerate into a trivial game. The considerations valid for the basic 2-agents model still hold in this case.

4. Discussion and concluding remarks

Recent contributions to the literature on climate change and sustainability, have stressed that the nature of such games might primarily be a coordinative one (FD2013, MS2018), rather than a prisoner's dilemma. Both efforts surely raised relevant issues. In particular, DF2013 pinpointed the need to achieve a full acceptance (by all players) of the guidance of scientific knowledge on the threat of the climate catastrophe as well as of the role of economic activity in triggering it, as an urgent pre-condition for allowing climate international negotiations to escape the trap of prisoner's dilemma. Similarly, though at the different scale of private investments in green vs brown technologies, MS2018 correctly pinpointed the need to urgently abate the uncertainty surrounding green investments as a pre-condition for coordination. However, as discussed in our review of these contributions (Section 2), previous results, though correct *per se*, are limited by the underlying hypotheses.

Consequently, in this note, after having supplied a definition of sustainability games, we departed from the same framework in MS2018 and, after having remarked that they implicitly postulated coordination so that their game no longer represents a true sustainability game, we expanded their framework to also include sustainability games. This allowed us to show that, as long as some degree of individual interest is involved, the intrinsic nature of the involved games will be that of a standard prisoner's dilemma. Consequently, the emergence of coordination will require - as well known from basic game theory – either external interventions or sufficiently far-sighted agents

capable to compare, over an extended time horizon, the short-term gains from individualistic behaviour with the long-term collective - and therefore also individual - welfare loss. The explicit inclusion of irreversible environmental degradation brings interesting insight into this.
Indeed, one could wrongly conclude that environment degradation, by lowering the defectors' payoff stream, just acts as a further discount term, thereby unavoidably forcing the emergence of cooperation. However, things are more complicated. Indeed, under conditions more easily holding when the rate of environment degradation is not too large, damages to the environment can frustrate cooperation rather than favouring it, making more likely that the final outcome of the game becomes a very long sequence of "deteriorated" games where coordination will only occur when the extent of the injury to Nature might be dramatic. The reason is that, under these conditions, environment degradation will sooner or later cause the game to lose its nature of prisoner's dilemma evolving into a trivial game with full green outcome as unique Nash equilibrium. At that stage, threatening to play brown forever will no longer be credible and coordination will fail. Based on the current projections on emissions and temperature growth (IPCC 2021) showing that a long span of time of increasing degradation rates might be observed before the currently planned interventions will allow to display significant results, the previous result suggests that many opportunities to coordinate right now i.e., in the moment when they return would be maximal, could be lost. This agrees with long-standing views on sustainability (Meadows et al., 2004).
Nonetheless, the dispute "prisoner's dilemma vs coordination" about the nature of sustainability games is not that trivial in view of the dramatically different policy prescriptions that would result. Dogmatic trust into coordination conveys the basically optimistic message that, besides the guidance of science stressed by DF2013 to solve coordination failures, everything could be accommodated by "laissez- faire". Obviously, a completely different attitude would be required under full acknowledgement that most sustainability games might be prisoner's dilemmas.
In view of the short time scales in the fight against the effects of climate change, the acknowledgement of the potential pervasiveness of the prisoner's dilemma is critically important especially given the dramatic extent of poverty and inequality at the global level, both in its current absolute magnitude and its increasing trend, as documented in a number of masterpieces (Stiglitz, 2012; Alvaredo et al 2018; Piketty, 2020). But inclusion of inequality in strategic interactions weakens the conditions to have a prisoner's dilemma and eventually can even prevent agents to play. This result, trivially holding in simple games, has been shown to hold under fairly general conditions (Wang et al 2009, Johnson and Smirnov 2018). **The detrimental role of inequality has been also confirmed in a number of recent studies in experimental economics (e.g., Tavoni et al 2011; Gross and Böhm 2020 and references therein).**
The obvious intuition in relation to sustainability games is that, under poverty conditions, "being brown" can be the only option available. In this regard, recent macrosimulation evidence (d'Alessandro et al, 2020) has shown that the benefits of the "green growth" scenario in achieving reductions in greenhouse gas emissions might pay the price of a parallel continuing increase in income inequality and unemployment. These phenomena would therefore worsen the conditions of sustainability games and/or increasing the proportions of the overall population that can only choose to be brown. From this perspective, the results reported here cannot but support their

major insight, namely the need for large scale social policies going parallel with mitigations interventions (d'Alessandro et al, 2020). [6]

This is especially true in the current moment, in view of the evidence that the COVID-19 pandemic has played a major redistributive role, first by more severely affecting more deprived population groups (e.g., Mena et al, 2020), but especially in enhancing inequalities in the future (e.g., Furceri et al 2021) even despite eventual achievement of full epidemic control.

To sum up, even though ability to enact coordination will be, no doubt, the key for a successful climate battle, nonetheless we believe that current undue emphasis on the coordinative nature of sustainability games might be deleterious in view of its implicit optimistic messages that could undermine attempts to enact those policies, first of all those aiming at inequality reduction, that would be a precondition for successful coordination. Again, borrowing from the top level, there is an endless list of instances ranging from (i) scientific debunking of climate change (Björnberg et al, 2017), (ii) the steady failure of climate agreements (Harris, 2007; Victor, 2012; Napoli 2012; Rosen, 2015; Clemencon, 2018) including the more or less systematic defection of the two major powers namely the US and China (in passing: currently disposing of the major research power and infrastructures), departing from president Trump's administration denialism about climate change (De Pryck and Gemenne, 2017), up to (iii) the recent statement (after pressure by environmentalist associations) of the German Federal Constitutional Court against the German government because of its insufficient actions (Federal Climate Change Act, 12 December 2019, Bundes-Klimaschutzgesetz – KSG) against climate targets and annual emission amounts as incompatible with fundamental human rights. All these are strong evidence of the persistence of defecting behaviour.

Acknowledgements. The authors thank two anonymous referees of the Journal whose valuable suggestions allowed to greatly improve the quality and exposition of the manuscript. Usual disclaimers apply.

References

Alvaredo, F., et al, 2018. The World Inequality Report. Harvard University Press.

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

Björnberg, K. E., Karlsson, M., Gilek, M., & Hansson, S. O. (2017). Climate and environmental science denial: A review of the scientific literature published in 1990–2015. Journal of Cleaner Production, 167, 229-241.

Cadsby, C.B., Maynes E., 1999. Voluntary provision of threshold public goods with continuous contributions: experimental evidence. Journal of Public Economics 71, 53–73.

Carraro C., Siniscalco D.,1993. Strategies for the international protection of the environment. J Public Econ 52:309–328.

[6] The previously cited studies in experimental economics (e.g., Tavoni et al 2011; Gross and Böhm 2020) also supply interesting indications for facing and reducing the effects of inequality and the tendency towards self-reliant behaviour, thereby promoting cooperation.

Carrozzo Magli, A., Della Posta P., Manfredi, P., 2021. The Tragedy of the Commons as a Prisoner's Dilemma. Its Relevance for Sustainability Games, Sustainability 2021, 13, x. https://doi.org/10.3390/xxxxx.
Clémencon, R., 2018. The Two Sides of the Paris Climate Agreement: Dismal Failure or Historic Breakthrough? Journal of Environment & Development, 25(1) 3–24.
D'Alessandro, S., Cieplinski A., Distefano T., Dittmer C., 2020. Feasible alternatives to green growth, Nature Sustainability volume 3, pages329–335 (2020).
DeCanio, S.J., Fremstad, A., 2013. Game theory and climate diplomacy, Ecological Economics 85, 177–187.
Delle, M., Jones, B., Olken, B.A., 2008. Climate Change and Economic Growth: Evidence from the Last Half Century. American Economic Journal: Macroeconomics 4.
De Pryck, K., & Gemenne, F. (2017). The denier-in-chief: Climate change, science and the election of Donald J. Trump. Law and Critique, 28(2), 119-126.
Furceri, D., et al., 2021. Will COVID-19 Have Long-Lasting Effects on Inequality? Evidence from Past Pandemics, International Monetary Fund reports.
Gross, J., & Böhm, R. (2020). Voluntary restrictions on self-reliance increase cooperation and mitigate wealth inequality. Proceedings of the National Academy of Sciences, 117(46), 29202-29211.
Hardin, G., 1968. The tragedy of the commons. Science 162, 1243–1248.
Harris P.G., 2007. Collective Action on Climate Change: The Logic of Regime Failure, Natural Resources Journal, 47, 195-224.
Heal, G., 2007. Discounting: a review of the basic economics. U. Chi. L. Rev., 74, 59.
Heitzig, J., Lessmann, K., Zou, Y., 2011. Self-enforcing strategies to deter free-riding in the climate change mitigation game and other repeated public good games. Proc.Natl. Acad. Sci. 108 (38), 15739–15744.
Heugues, M., 2013. The global emission game: on the impact of strategic interactions between countries on the existence and the properties of Nash equilibria, Milano: Fondazione Eni Enrico Mattei. Available at: http://www.feem.it/userfiles/attach/2014181222104NDL2013-108.pdf, Accessed date: 4 December 2020.
Johnson, T., Smirnov, O., 2018. Inequality as information: Wealth homophily facilitates the evolution of cooperation. Scientific Reports, 8(1), 1-10.
Inter-governmental Panel on Climate Change, IPCC, 2021. Sixth Assessment Report, https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf
Meadows, D.H., Randers, J., Meadows, D.L., 2004. The limits to growth: the 30-year update, the MIT press.
Mena, G. E., et al, 2021. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. Science, 372(6545).
Mielke, J., Steudle, G.,A., 2018. Green Investment and Coordination Failure: An Investors' Perspective. Ecological Economics 150, 88–95.
Napoli, C., 2012, Understanding Kyoto's Failure, SAIS Review of International Affairs, Johns Hopkins, 2, pp. 183-196, 10.1353/sais.2012.0033.

Nordhaus, W., 2015. Climate Clubs: Overcoming Free-riding in International Climate Policy, American Economic Review 2015, 105(4): 1339–1370, http://dx.doi.org/10.1257/aer.150000011339
Pigou, A.C., 1920. *The Economics of Welfare.* Macmillan, London.
Polasky, S., et al., 2019. Role of economics in analyzing the environment and sustainable development. Proceedings of the National Academy of Sciences, 116(12), 5233-5238.
Pindyck, R.S., 2012. Uncertain outcomes and climate change policy. Journal of Environmental Economics and Management 63(3), pp. 289-303.
Piketty T., 2020 Capital et idèologie, Ed du Seuil, Paris.
Stiglitz, J.E., 2012. The price of inequality: How today's divided society endangers our future. WW Norton & Company.
Rapoport, A., 1988. Experiments with N-Person Social Traps II: Tragedy of the Commons. J. Conflict Resolution., 32, 473-488.
Rosen, A. M. (2015). The wrong solution at the right time: The failure of the kyoto protocol on climate change. Politics & Policy, 43(1), 30-58.
Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. Proceedings of the National Academy of Sciences, 108(29), 11825-11829.
Victor, D.G., 2011. The collapse of the Kyoto Protocol and the struggle to slow global warming. Princeton University Press.
Wang, J., Fu, F., Wu, T., Wang, L., 2009. Emergence of social cooperation in threshold public goods games with collective risk, PHYSICAL REVIEW E 80, 016101.

Appendix: details on the multi-agent game results

In this appendix we provide a few more details on the analysis of the one-shot multi-agent game reported in section 3.2. A standard analysis leads to the following conditions for the different outcomes of the game.

- $\frac{1-2\beta}{2c} > \frac{2\gamma+1}{2d} \rightarrow \frac{d}{c} > \frac{2\gamma+1}{1-2\beta}$, the game is trivial: the brown configuration is the only equilibrium and is preferred by all agents with respect to a full green configuration (this possibility is listed just for the sake of completeness);
- $\frac{(n+\gamma)^2}{2dn^2} - \frac{\beta(n-\beta)(n-1)}{cn^2} > \frac{1-2\beta}{2c} \rightarrow \frac{d}{c} < \frac{(n+\gamma)^2}{n^2+2\beta^2-2\beta n(1+\beta)} \underset{n\to\infty}{\longrightarrow} \frac{d}{c} < 1$, the game is trivial: the green configuration is the only equilibrium (this possibility is listed just for the sake of completeness);
- $\frac{1}{2c} < \frac{2\gamma+1}{2d} \rightarrow \frac{d}{c} < 2\gamma + 1$, the game is a coordination game;
- $\frac{1}{2c} > \frac{2\gamma+1}{2d} \left(> \frac{1-2\beta}{2c}\right) \rightarrow \frac{d}{c} > 1 + 2\gamma$, the game is a prisoner's dilemma.

The previous conditions identify the relevant thresholds reported in the main text for the case of a large number of agents, namely $T_G = 1, T_C = 1 + 2\gamma$ and

$$T_P = \frac{1+2\gamma}{1-2\beta} = T_P \cdot \varepsilon(\beta),$$

where $\varepsilon(\beta)$ is increasing in β. Notably, $\frac{\beta}{n}$'s impact on T_C when n goes to infinity. The environmental damage β would not disappear if one reinterpreted the damage function of the brown technology as depending on the absolute number of brown players e.g., as $-\beta n$ instead of following MS2018's setup based on relative proportions only. All other results would not be affected.