Climate policy in need of plan B

Prepared for a Festschrift in honor of Philippe Aghion and Peter Howitt

Christer Fuglesang[a.] and John Hassler[b.]

a. Space Centre, KTH Royal Institute of Technology,
b. IIES, Stockholm University

Abstract

The chapter first argues that a transition to global climate neutrality by 2050 is a good plan A for dealing with climate change. If the plan is built on carbon pricing and has the necessary complementary policies, it need not compromise economic growth and distributional concerns. Second, it argues that due to the large uncertainty about the consequences of emitting carbon dioxide, one or several plan B's should also be developed. A plan B should be executed in case climate sensitivity is much higher than expected or international climate policy coordination breaks down. A model is constructed to analyze the claim that developing such a plan B might endanger the successful implementation of plan A. The analysis reveals that the argument is logically correct, but it is argued that the insurance value of having a plan B likely exceed the negative effects of developing one. Third, the paper, describes a possible plan B in the form sunshades in space to reduce the inflow of sunlight to Earth, thus mitigating the greenhouse effect

1 Introduction

Climate change due to emissions of greenhouse gases is by some portrayed as an acute threat to our civilization. Emissions of CO_2 and other greenhouse gases strongly affect the climate in a non-linear way with tipping points and irreversibilities. Others argue that even though greenhouse gases do affect the balance of incoming energy from the Sun and the outgoing from Earth to space, the consequences of this for climate as well as the consequences of climate change for human welfare are exaggerated in media and the political discussion. Unfortunately, science has not yet been able to credibly verify were in this large interval truth lies. Climate policy, or the lack of it, must then be chosen in a situation with very large uncertainty.

Climate change is driven by global aggregate emissions but there is no single entity that decides their level. Instead, emissions are caused by the decisions of billions of individuals, firms and other agents. Policies that affect these decisions are decided by hundreds of individual governments with access only to weak and imperfect ways of solving coordination and commitment problems. At the same time, it is clear that cooperation and commitment is necessary to deal with climate change.

These complications must be born in mind when giving scientific policy advice. Deriving optimal policy in a model where risks can be measured with objective probabilities and policy is chosen by a benevolent central planner is not sufficient if these conditions are not satisfied in reality. The area of climate change is an example of this.

In this chapter, we argue in favor of an orderly transition to climate neutrality towards the mid of the current century. The forces of creative destruction are strong and can transform society completely. By using carbon pricing these forces can be used to steer technical change in a direction making such a transition possible without large costs. Continued growth and catch-up in the developing countries of the world would not be sacrificed. If the transition is allowed to take a few centuries, it does not cost much also if it turns out that the climate change sceptics are right. If they are wrong, the transition would be highly beneficial. Thus, a transition to climate neutrality is a good insurance policy. Given that it is not costly, coordination and commitment problems are at least partially mitigated. This is what we will call plan A.

However, we will also argue that plan A should be complemented by the development of a plan B. This is due to the large uncertainty about how sensitive climate is to emissions and how sensitive human welfare is to climate change. Likely, plan A is sufficient but we cannot rule out that it is not. Due to very high costs and the associated political impossibility, a quick stop for using fossil fuel is not a viable plan B.

Instead, we argue that other solutions need to be sought.

In a choice situation, like that of a benevolent central planner choosing policy to maximize welfare, it can only be good to have more options. However, in the real-world environment where climate policy is determined, this is not necessarily the case. Having access to a plan B may make the outcome worse. We will investigate this argument in a formal model and show it to be relevant in the case of climate policies. However, we will argue that the argument is likely not strong enough to overturn the benefits of a plan B. Therefore, we will end the chapter by describing one of potentially many possible plan B – sunshades in space.

2 Climate change uncertainty and insurance values

2.1 Uncertainty about climate sensitivity to emissions

Higher concentration of greenhouse gases reduces the outflow of energy from Earth. Given a constant inflow, this creates a surplus in Earth's energy balance that leads to higher ground temperatures. Since a higher temperature increases the outflow, a balanced energy budget will eventually be restored, but at a higher temperature. There are several greenhouse gases and in principle all gases with molecules having more than two atoms produce greenhouse effects. Carbon dioxide, CO_2, is special in the sense that it is emitted by human activities in quantities that make it the by far largest contributor to the surplus in the global energy budget. Furthermore, it is special in the sense that a substantial share of emissions stay in the atmosphere for thousands of years.[1]

The fundamental principles behind the warming effect of greenhouse gases have been known for centuries and the effect of a higher atmospheric CO_2 concentration were quantified already in the 19th century by Arrhenius (1896). His empirical analysis concluded that a doubling of the CO_2 concentration increases the average temperature at the equator by 5°C and by 6°C at latitude 60 (e.g., Stockholm and Anchorage). A few years later, he reduced the estimate of this so called *equilibrium climate sensitivity* to a global increase of 4°C (Lapenis, 1998). This number is still within the range of uncertainty provided by the latest sixth IPCC report (IPCC, 2021), which states a 90% confidence interval between 2 and 5°C and a best guess of 3 degrees. These confidence intervals have not changed much since the first IPCC report, that stated an interval between 1.5 and 4.5°C and a best estimate of 2.5°C (IPCC, 1990).[2]

The later reports also provide what is called *likely,* specified as 67%, confidence intervals. In the fourth report (IPCC, 2007) this is 2 to 4.5°C with a best guess of 3°C. In the fifth (IPCC, 2013), the interval is slightly wider, 1.5 to 4.5°C and no best guess is provided. In the sixth report (IPCC, 2021) the likely interval is narrowed somewhat to 2.5 to 4°C and the best guess of 3°C is reintroduced.

These ranges are large and have quite substantial impacts on how much emissions in the future can be allowed given a particular target for the global mean temperature. This can easily be appreciated by using the finding that advanced circulation models that include also a descrip-

[1] This is not the case for methane, which is a very potent greenhouse gas and the second largest contributor to the energy budget surplus, but that disolves in the atmosphere in a few decades.

[2] The first report does not provide significance levels for the confidedence interval.

tion of the circulation of carbon between the different reservoirs (mainly the atmosphere, the biosphere and the oceans) produce the result that global warming at a particular point in time is proportional to the historically accumulated amount of emissions up to that point. This finding was first noted by Matthews et al. (2009), and is now incorporated in later IPCC reports. The sixth report (IPCC, 2021) states "*This Report reaffirms with high confidence the AR5 finding that there is a near-linear relationship between cumulative anthropogenic CO_2* emissions and the global warming they cause. Each 1000 $GtCO_2$ *of cumulative CO_2 emissions is assessed to likely cause a $0.27°C$ to $0.63°C$ increase in global surface temperature with a best estimate of $0.45°C$.*" The quantitatively most important uncertainty producing this range is regarding climate's sensitivity to CO_2 concentrations.

Using IPCC's range for the emission sensitivity and the estimate that current accumulated emissions are 2400 $GtCO_2$ (IPCC, 2021), we can easily compute that our previous emissions imply a global warming commitment of between 0.65 and 1.5°C.[3] Suppose now that we set the target to 2 degrees and hope that the lower value of the climate sensitivity is the correct one. Then, we can emit a further 5000 $GtCO_2$, i.e., twice as much as we have emitted so far in history. If it then turns out that our hope were in vain, and instead the emission sensitivity is at the upper end of the range, the temperature would rise to 4.7 degrees. These calculations uses the *likely* range, which means a 67% likelihood. Using the same argument for the wider 90% interval, or including also less likely but possible climate sensitivities, even more striking consequences of making policy errors would result. Below, we will return to the economic consequences of such errors.

2.2 Damage sensitivity to climate change

Climate policy also must take into account that the effects on human welfare of climate change is highly uncertain. Nordhaus and Moffat (2017) and Howard and Sterner (2017) provide metastudies of the global aggregate effect of climate change on humans, where effects are monetized and expressed relative to GDP. Used on existing studies, Nordhaus and Moffat (2017), provide an estimated range of damages between -0.1 and 3.4% of global GDP for $3°C$ global warming. The studies underlying Howard and Sterner (2017) have a range from 0 to 12% at the same temperature.

Hassler at al. (2018), uses IPCC's likely ranges for the climate sensi-

[3] These estimates are all consistent with the current temperature increase of $1.09°C$, since they are based on different assessment of temporary factors affecting the current temperature.

tivity and the variation in damage sensitivities in Nordhaus and Moffat (2017) to show that uncertainty in terms of the difference in policy implications is very large. In particular, they show that if climate sensitivity and damage sensitivity is at the lower end of the likely ranges, the optimal tax is practically zero at 1.9 US$/ton CO_2. If both sensitivities are high, it is instead 72.2 US$/ton CO_2.[4] In both cases, the tax rate should increase at the rate of global GDP growth. The numbers are of course conditional on a number of other parameters, e.g., the rate at which future welfare is discounted. A lower discount rate increases the optimal tax rate but does it approximately proportionally regardless of sensitivities so that the ratio between the high and the low tax rate is maintained. Reducing the discount rate from 1.5% to 0.1%, as suggested e.g., in the Stern Review, increases all optimal taxes approximately by a factor nine.

Hassler et al. (2018) also show that the two types of policy uncertainty, about climate sensitivity and about damage sensitivity, are of similar magnitude in terms of optimal tax implications. A high climate sensitivity and a low damage sensitivity implies an optimal tax of 10.0 and the converse combination an optimal tax of 12.4 US$/ton CO_2.

2.3 Climate Policy under uncertainty

Characterizing optimal policy in situations with risk, i.e., when (objective) probabilities can be assigned to different possible states of the world, is standard since long in economics. Cost-benefit analysis of risky policy making is highly useful and practiced in government agencies all over the world, also when dealing with delicate issues about e.g., health, life and death. The discussion above, where probabilities and probability intervals from IPCC were used, suggests that the same approach could be used when deriving recommendations for climate policy. However, we argue that the probabilities stated by IPCC are not well suited for being used as input in a model where optimal policy is derived by finding the policy that maximizes expected utility. First, the probabilities only cover a fairly small range of the set of possibilities, failing to assign probabilities to unlikely but possible states of the world. As suggested by Weitzman (2009), such extreme possibilities may be highly relevant for what is good policy. Second, the probabilities stated by IPCC are quite judgemental. This is inevitable in a situation where different models are consistent with observed data while producing highly different predictions for the future. These two arguments make the probabili-

[4] Sometimes, optimal taxes and emissions are expressed in massunits of carbon. Since one massunit of carbon produces 3.66 units of CO_2, conversion is straightforward.

ties stated by IPCC less suitable to use as input in the calculation of expected utility.

A high degree of uncertainty, *ceteris paribus,* increases the value of waiting (Pindyck, 1991). One could then make the argument that action should be delayed. However, this conclusions is likely wrong in the case of climate policy. There is no value of waiting in itself. Instead the value arises due to an expected flow of information. By waiting, a more informed and thus better decision can be made. However, based on what we so far have observed, it seems unlikely that the uncertainty about the relevant sensitivities will be much reduced in the near future. Thus, the value of waiting is not high.

Integrated assessment models building on the seminal work by Nordhaus (see KVA, 2018 for a summary) describe the interaction between the climate and the economy. They are global and long-run to be able to deal with the determinants and consequences of climate change and climate policy. Despite some criticism (see e.g., Pindyck, 2013), we argue that integrated assessment models are highly useful also in situations of large uncertainty about key parameters.

Hassler et al. (2018) shows the usefulness of integrated assessment models in a simple exercise. They use IPCC's likely range for the climate sensitivity and the damage sensitivity mentioned above and consider the consequences of possible policy mistakes. In a situation with a high degree of uncertainty, policy mistakes are likely. What are then the consequences of setting a low carbon tax, hoping for sensitivities at the lower end of the likely range, if this turns out to be the wrong choice? Conversely, what happens if a high tax is set, while it turns out that a low was optimal? The result in the paper is that the consequences of these policy mistakes are highly asymmetric. This is shown in Figure 1. As we see, the consequences of in vain setting a high tax rate are very small compared to the opposite policy mistake. It could have been the case that the costs of policy mistakes are high and symmetric. Then, the problem would be of the *wicked* nature where policy advice is very difficult to give. Instead, the asymmetry points to an ambitious climate policy being a good insurance – fairly cheap also if not needed *ex post*, but good to have in case. This is a quite policy-relevant result that relies on the use of integrated assessment models.

In contrast to Rudik (2020), the result in Hassler et al. (2018) is derived in a model without learning. However, learning might as discussed above not be sufficiently fast to overturn the results. Furthermore, the cost of a transition to climate neutrality depends crucially on how fast the transition is. A fast transition is much more costly and waiting might lead to such a transition becoming necessary. This strengthens the case

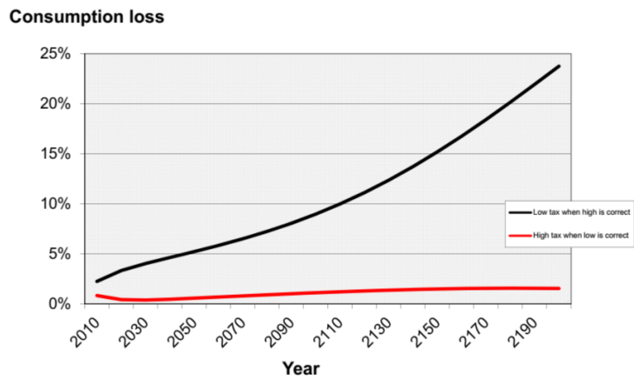

Figure 1: Asymmeric policy consequences of policy mistakes.

for not waiting, but underscores the possibility that policy will ex post turn out to be suboptimal.

A fast transition is costly since the economy is not very flexible in the short run. Hassler et al. (2021) show that in the short run, a reduction of fossil fuel consumption can basically only be done by reducing production. This since the production function in the short run requires energy in proportions almost fixed to output. In the longer run, however, the elasticity between energy and other inputs is close to unity, allowing output to grow without energy use increasing. The mechanism behind this is directed technical change – as the price of energy increases, technological effort is redirected towards making production more energy efficient. Hassler et al. (2021) use aggregate evidence to draw these conclusions. See Aghion et al. (2016) for microeconomic evidence in the same direction. Building on the same ideas, but using a less stylized model where a number of short-run frictions are taken into account, IMF (2020) also shows that a global transition to climate neutrality over three decades based on carbon pricing need not be costly in terms of lost output if complemented by some initial stimulus policy to counteract the initial contractionary effects of carbon prices. The pricing policy shold be complemented by subsidies to green technology development. This provides an important complement to make the transition easier, but is no substitute for pricing. Furthermore, the revenues from pricing emissions are more than sufficient to undo unintended distributional effects (IMF, 2020). The fact that at least in the OECD, the problem of too low or absent carbon pricing does not apply to transportation fuel (OECD, 2016), means that that difficulties related to public disent like that of the French yellow wests may be smaller than sometimes claimed. Of course, many political hurdles before a global agreement on carbon pricing can be achieved. This is not a paper about these, but we want to note that the recent aggreement between the G7 countries to aim for the implementation of a global minimum tax rate on corporate revenues indicates that such agreements are possible.

The conclusion of this section is therefore that an orderly transition to climate neutrality around 2050 based on carbon pricing is a good insurance against the uncertain consequences of using fossil fuel. This is *likely* going to keep climate damages at an acceptable level and the insurance premium of this policy is not too high. This should therefore be plan A to deal with climate change. A key limitation, however, is captured by the word *likely*. The discussion above has used the *likely* range of uncertainty, by IPCC described as a 67% confidence interval. But climate and damage sensitivities may be substantially higher and plan A might also fail due to international policy cooperation breaking down. We therefore argue that we also need to consider the development of a plan B, for unlikely but possible really bad scenarios.

There are arguments against developing a plan B and the idea is even criticized as being highly dangerous. The arguments behind such claims are usually not formal, an exception being the working paper by Acemoglu and Rafey (2019). IPCC (2014) argues that plan B's like geoengineering may be seen as a substitute for plan A and reduce the effort in it's implementation which may be highly disadvantageous. In the next section, we will examine this argument in a formal but simple game-theoretic model.

3 The cost and benefit of a climate plan B

3.1 An emission abatement game

In this section, we describe a game where agents have a common interest in reducing climate change damages through emission abatement. As in reality, a key complication is the free-rider problem, implying that there is an incentive to deviate from the cooperative solution that maximizes joint welfare by emitting more. In order to maintain the cooperative solution, each player has access to a way of reducing the incentive for other players to deviate from the cooperative low-emission solution. However, this requires a costly investment in early phases of the game and a stronger incentive reduction requires a higher investment. Without any such investments, the incentive to deviate and choose high emission levels will prevail making the cooperative solution impossible. We do not specify the concrete character of these investments, but it may help to think about them as time consuming participation in international negotiations as well as building up the capacity to punish other players by e.g., imposing trade barriers.

The purpose of constructing the model is to analyze the costs and benefits of developing a plan B against climate change. With such a plan, we mean the development of a technology that can drastically reduce the negative consequences of climate change. Developing the technology is assumed to have negligible costs but using it imposes costs on society. These costs stem both from direct technological costs and from that it may have negative side effects. Importantly, we assume that these costs are independent of the amount of climate damages while the direct benefits of executing plan B increase in the amount of climate damages that are mitigated by it. Thus, execution of plan B is beneficial only if climate damages are sufficiently large. As we will see, however, the mere development of the plan has consequences for the game that induces costs on society also in the case when the plan is not executed. Specifically, if the plan is developed, more needs to be invested in maintaining the low-emission cooperative solution in the cases when this is socially preferable over plan B. The existence of a developed plan B also opens up for multiple equilibria where one is plan B despite it not beeing socially preferable.

Let us now describe the game in more detail with the help of Figure 2. The game is a three-stage game with two players A and B. The *first stage* starts with player A deciding whether or not to develop plan B.[5] It ends with nature choosing the sensitivity of climate damages to

[5] This asymmetry is innoccous since there are no interests of conflict at this stage.

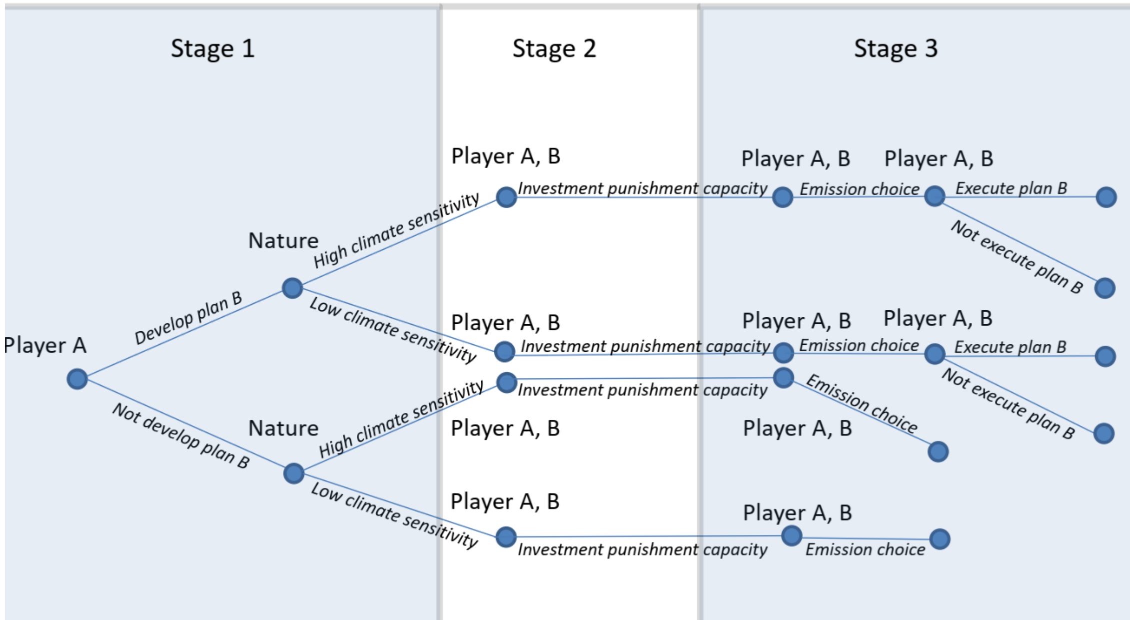

Figure 2: A climate abatement game with a plan B.

emissions. This sensitivity incorporates both the sensitivity of climate to emissions and the sensitivity of damages to climate change. We call this the *damages sensitivity* (to emissions) and label it γ. For simplicity, we assume that it is either strong, γ_S or weak, γ_W. The probability of strong damage sensitivity is π_{γ_S}.

In the *second stage*, both players simultaneously decide on investing in the capacity to later punish their opponent for deviations from the low emission cooperative solution knowing both γ and whether a plan B has been developed. For players $i \in \{A, B\}$, the punishment capacity is denoted Π_i and building it has costs given by the increasing function $c(\Pi_i)$ with $c(0) = 0$.

In the final *third stage*, players first choose the amount of emissions simultaneously. For simplicity, we assume that the choice is restricted to either low or high emissions. Choosing high emissions has a private benefit to the player denoted e. High emissions will also increase climate damages for both players but we will consider the realistic case when these private benefits are larger than the increase in climate damages holding the actions of the other player fixed. Any player i who chooses high emissions will also face the punishment Π_{-i} determined by the other player in the previous period. These emission levels will lead to symmetric climate damages (if plan B is not executed) that depend on γ and the aggregate amount of emissions. The damages are denoted $D_{LL}(\gamma)$, $D_{LH}(\gamma)$ and $D_{HH}(\gamma)$ for the three possible aggregate emission levels (both choosing low, one low and one high and both high). We assume $D_{LL}(\gamma) < D_{LH}(\gamma) < D_{HH}(\gamma)$ and for simplicity that $D_{LH}(\gamma) - D_{LL}(\gamma) = D_{LH}(\gamma) - D_{HH}(\gamma)$. If no plan B was developed in stage 1, this concludes the game. If plan B was developed, any of the agents can decide to execute it. If it is executed, any climate damage is neutralized and replaced by a fixed cost denoted p. Here, there will be no disagreement between the players and the plan will be executed iff climate damages are higher than p.

To make the analysis interesting, we assume that $D_{LL}(\gamma_W) < p$ while both $D_{LH}(\gamma_W) > p$ and $D_{LL}(\gamma_S) > p$. Thus, plan B will be executed if it is developed unless damages sensitivity is weak and the cooperative solution with both players choosing low emission levels materializes. Furthermore, we assume that $e > D_{LH}(\gamma) - D_{LL}(\gamma)$ so that choosing high emissions is privately optimal (a dominating strategy) if $\Pi_{-i} = 0$.

3.2 Equilibrium

We solve for the equilibrium by backward induction.

3.2.1 Stage 3

Consider first the case when no plan B has been developed. In stage 3, the game has the following structure:

Stage 3 game - no plan B developed

		Player B	
		Low	*High*
Player A	*Low*	$-D_{LL}(\gamma), -D_{LL}(\gamma)$	$-D_{LH}(\gamma_L), e - D_{LH}(\gamma) - \Pi_A$
	High	$e - D_{LH}(\gamma) - \Pi_B, -D_{LH}(\gamma)$	$-D_{HH}(\gamma) - \Pi_B, D_{HH}(\gamma) - \Pi_A$

Under the assumptions given above, cooperation with both players choosing low emission is the only Nash equilibrium iff

$$\Pi_{A,B} \geq e - (D_{LH}(\gamma) - D_{LL}(\gamma)) \equiv \bar{\Pi}(\gamma). \tag{1}$$

In words, the punishment must be large enough to deter the value of deviation, which is given by the direct value e and the difference in climate damages under cooperation and deviation.[6] If the punishments chosen by one player is lower than $\bar{\Pi}(\gamma)$, emissions of the other player will be high.

Now consider the case when plan B was developed. The game outcome now depends on whether γ is weak or strong. In the case when

[6]We assume that when indifferent, players choose low emissions.

$\gamma = \gamma_W$, plan B will be executed if and only if at least one player choose high emission levels. The game is then as given below.

Stage 3 game - plan B developed, weak damage sensitivity

		Player B	
		Low	*High*
Player A	*Low*	$-D_{LL}(\gamma_W), -D_{LL}(\gamma_W)$	$-p, e-p-\Pi_A$
	High	$e-p-\Pi_B, -p$	$e-p-\Pi_B, e-p-\Pi_A$

The cooperative equilibrium with low emissions then requires

$$\Pi_{A,B} \geq e - (p - D_{LL}(\gamma_W)) \tag{2}$$

Now, since we have assumed that $p < D_{LH}(\gamma_W)$, the temptation to deviate is larger than if plan B was not developed (the RHS of (2) is larger than the RHS of (1). Thus, punishment must be larger to sustain cooperation.

Furthermore, if player i chooses high emissions, the value for the other player $-i$ of also doing it will be e since plan B will be executed whatever the choice of player $-i$. Thus, $\Pi_{-i} = e - (p - D_{LL}(\gamma_W))$ does not rule out the high emission outcome. To do that, the punishment for deviation must be

$$\Pi_{A,B} \geq e.$$

To limit the number of cases and to stack the case against a plan B, we assume that in the case of multiple equilibria, i.e., when $\Pi_i \in [e - (p - D_{LL}(\gamma_W)), e]$ the high emissions outcome results.

The final case to consider in stage 3 is when plan B was developed and damage sensitivity is strong. Then, plan B will be executed regardless of emission choices and the game is now:

Stage 3 game - plan B developed, strong damage sensitivity

		Player B	
		Low	*High*
Player A	*Low*	$-p, -p$	$-p, e-p-\Pi_A$
	High	$e-p-\Pi_B, -p$	$e-p-\Pi_B, e-p-\Pi_A$

Here, the implementation of plan B removes the free-rider problem. Provided $e > \Pi_{-i}$,it is is a dominating strategy to choose high emissions and otherwise emissions are low. Payoffs to players will then be $e-p-\Pi_{-i}$ in the former case and $-p$ in the latter.

This concludes the analysis of stage 3.

3.2.2 Stage 2

Now, consider stage 2 in which the investment in punishment capacity is done simultaneously given nature's choice of damage sensitivity γ and whether or not plan B has been developed. Without loss of generality, we can restrict the policy space to choosing $\Pi_i \in \{0, \bar{\Pi}(\gamma), \}$ if no plan B is developed and $\Pi_i \in \{0, e\}$ if it is.

Consider first the case when plan B has not been developed. Incorporating the equilibrium in stage 3 yields the following game structure in stage 1:

Stage 2 game - no plan B developed

			B
		$\Pi_B = 0$	$\Pi_B = \bar{\Pi}(\gamma$
A	$\Pi_A = 0$	$e - D_{HH}(\gamma), e - D_{HH}(\gamma)$	$-D_{LH}(\gamma), e - D_{LH}(\gamma) - c(\bar{\Pi}(\gamma$
	$\Pi_A = \bar{\Pi}(\gamma)$	$e - D_{LH}(\gamma) - c(\bar{\Pi}(\gamma)), -D_{LH}(\gamma)$	$-D_{LL}(\gamma) - c(\bar{\Pi}(\gamma)), -D_{LL}(\gamma) - c(\bar{\Pi}(\gamma$

Here, cooperation is the unique outcome if

$$c\left(\bar{\Pi}(\gamma)\right) < D_{LH}(\gamma) - D_{LL}(\gamma). \tag{3}$$

This condition implies that it is worthwhile to take the cost and effort to induce the cooperative low emission equilibrium. To make the analysis interesting, we assume this to be the case regardless of the damage sensitivity.

Now, consider the case when plan B has been developed. The game now depends on whether damage sensitivity is weak or strong. Consider first the case of low climate sensitivity. The game is then:

Stage 2 game - plan B developed, weak damage sensitivity

	$\Pi_B = 0$	$\Pi_B = e$
$\Pi_A = 0$	$e - p, e - p$	$-p, e - p - c(e)$
$\Pi_A = e$	$e - p - c(e), -p$	$-D_{LL}(\gamma_W) - c(e), -D_{LL}(\gamma_W) - c(e)$

Here, we see that developing plan B implies that no cooperation where $\Pi_{A,B} = 0$ always is an equilibrium. Additionally, $\Pi_{A,B} = e$ is also an equilibrium if

$$c(e) < p - D_{LL}(\gamma_W). \tag{4}$$

In words, the cost of creating the capacity to punish at a level e, which is $c(e)$, must be lower than the benefit of creating that capacity, which is $p - D_{LL}(\gamma_W)$. Thus, the development of a plan B makes cooperation more fragile, subject to multiple equilibria. We assume this multiplicity is resolved by a stochastic mechanism implying that the cooperative low-emission equilibrium results with a probability π_C.

In addition, since $c(e) > c(\bar{\Pi})$ the cost of implementing the low emission equilibrium is higher and since also $p < D_{LH}(\gamma_W)$ the condition for the existence of the cooperation equilibrium is tighter. Here, we focus on the case when $c(e) < p - D_{LL}(\gamma_W)$ so that also when plan B has been developed, it is worth the effort to induce low emissions and no execution of plan B if the damage sensitivity is weak.

Finally, consider the case when damage sensitivity is strong, in which case plan B will always be executed. The game is:

Stage 2 game - plan B developed, strong damage sensitivity

	$\Pi_B = 0$	$\Pi_B = e$
$\Pi_A = 0$	$e - p, e - p$	$-p, e - p - c(e)$
$\Pi_A = e$	$e - p - c(e), -p$	$-p - c(e), -p - c(e)$

Here, the only equilibrium is $\Pi_{A,B} = 0$. This concludes the analysis of stage 2.

3.2.3 Stage 1

In stage 1, the decision of whether or not to develop plan B is taken by player A before she knows the realization of the damage sensitivity. If plan B is not executed, the payoff is

$$\underbrace{-\left(1-\pi_{\gamma_S}\right)\left(D_{LL}(\gamma_W) + c\left(\bar{\Pi}(\gamma_W)\right)\right)}_{A_1} + \underbrace{-\pi_{\gamma_S}\left(D_{LL}(\gamma_S) + c\left(\bar{\Pi}(\gamma_S)\right)\right)}_{A_2}$$

If plan B is developed, the payoff is instead

$$-\underbrace{\left(1-\pi_{\gamma_S}\right)\left(\pi_C\left(D_{LL}(\gamma_W) + c(e)\right) + (1-\pi_C)(e-p)\right)}_{B_1} + \underbrace{-\pi_{\gamma_S}(e-p)}_{B_2}.$$

The term A_1 is the payoff if damage sensitivity is weak and no plan B is developed. It is higher than the corresponding term B_1 when plan B was developed for two reasons. First, it takes higher effort to implement the low emissions equilibrium $\left(c(e) > c\left(\bar{\Pi}_N(\gamma_W)\right)\right)$ and it may happen that plan B is executed even if the low emissions equilibrium is socially preferable $\left(e - p > D_{LL}(\gamma_W) + c\left(\bar{\Pi}_N(\gamma_W)\right)\right)$. On the other hand, the term A_2 is smaller than B_2 since if damage sensitivity is strong, implementing plan B is better than also the low-emissions equilibrium.

Note that in this stage of the game, there are no coordination issues. The choice of any of the player coincides with what they jointly would choose.

From the analysis above follows that:

Conclusion 1 *Under the assumptions in this section, there is a* $\bar{\pi}_{\gamma_S} > 0$ *such that iff* $\pi_{\gamma_S} \geq \bar{\pi}_{\gamma_S}$*, a plan B will be developed in stage 1. This is socially beneficial.*

Here, it is important to note two things. First, if and only if the probability of a high damage sensitivity is sufficiently high, plan B should be developed. Second, the threshold probability is strictly higher than zero. The latter is due to the finding that the existence of a plan B has negative effects on the outcome of the game if plan B is not needed. Specifically, it increases the cost of implementing plan A in case the damage sensitivity is weak and it may be implemented also when not needed from a social perspective. Thus, a plan B should not be developed unless the probability that it is socially benefical is sufficiently much above zero. This contrasts to the case when there is no coordination game. Then developing a plan B at zero cost can never reduce welfare.

4 Sunshades in space – a potential plan B

We now turn to describing a plan B against climate change, namely reducing the amount of incoming sunlight to Earth. There are several ways of doing this. The perhaps most well-known method is to inject aerosols into the stratosphere creating a cooling effect similar to the one experienced after large volcanic eruptions. A substantial amount of research has been done on this method including work by economists (see e.g., Smith, 2020, Smith and Wagner, 2018 and Wagner & Weitzman, 2015).

A less well-known method is to use shades in space. This was until recently considered impossible due to prohibitively large costs of rocket launches. When launching rockets were the unique privilege of state agencies like NASA, launch costs used were in the order of ten million dollar per ton. Costs have, however, already come down by almost an order of magnitude. Private initiatives by Elon Musk and others to create more cost efficient and reusable rockets point to the possibility to reduce the cost further by a least one, perhaps even two orders of magnitude. Given this development, Fuglesang and Garcia de Herreros Miciano (2021) recently constructed a concrete proposal for a plan B based on sunshades. We use the remainder of this section to describe the main ideas there.

4.1 Where to place the sunshades

A first question is where the sunshades should be located. One possibility is to have them circulating around Earth like satellites. However, this solution has several disadvantages. One is that this part of space is scarce due to the need for communication and observation satellites. More congestion here obviously increase collision risks. Furthermore, sunshades this close to Earth would be visible which has potentially negative side-effects and since they are circulating Earth, they would not shade any incoming sunlight when being themselves shaded by Earth.

A much better position is at a so called Lagrange point where the gravitational forces of Earth and the Sun together with the centripetal force of rotation around the Sun, balance each other. This occurs at a point around 1 percent away from Earth towards the Sun or around 4 times the distance to the Moon. An object placed at the Lagrange point would stay there unless affected by other forces. This location is much less crowded than low Earth orbits. Furthermore, sunshades placed there would always be in zenith and so far away that they would not be visible to the naked eye.

In reality, however, other forces than the gravitational would affect a sunshade placed at the Lagrange point. In particular, the solar radiation pressure (the force created by photons from the Sun bouncing on the shade) needs to be taken into account. To balance also this force, the shades needs to be placed closer to the Sun so that the increased gravitational force balances the solar radiation pressure. How much closer depends on the mass and the reflectivity of the shade. A lower mass and a higher reflectivity implies that the shade needs to be put closer to the Sun, thus further away from Earth. A larger distance to Earth has the negative consequence that the shade needs to be larger to decrease the same amount of sunlight reaching Earth.

Reflectivity should be minimized to minimize the distance between Earth and shades.[7] A higher mass is obviously costly in terms of launch costs. It turns out that to minimize the mass the shades should be put a distance 1.58% of the distance to the sun. This point is actually independent of reflectivity and the share of sunlight to be shaded, however the amount of the mass there increases linearly with both reflectivity and amount shading. This translates into an optimal areal density (mass per m^2) for the whole sunshade spacecraft, which depends on the reflectivity. Given already existing technologies for sunsails, it is not difficult to create sunshades with a mass equal to that optimum.

4.2 How large area needs to be shaded?

The size of the sunshades of course depends on how large cooling effect is required. It is convenient to express the latter in terms of how

[7] The theoretical optimum would be to have the shades absorb all energy from the Sun and transmit it as infrared radiation towards Earth, where it would be reflected by the atmosphere.

large reduction in energy inflow is wanted. The average (over time and space) inflow of energy to Earth from the Sun is 340 W/m^2. The latest IPCC report (IPCC 6, 2021) reports that the greenhouse effect of CO_2 has reduced outflow by 2 W/m^2 compared to 1850. In the same time period the CO_2 concentration has increased by 44%, from 286 to 410 ppm. Other greenhouse gases, methane and N_2O in particular, have non-negligible effects (0.5 and 0.2 W/m^2) and also leave the atmosphere much faster. Further, IPCC reports the estimated effect of the energy outflow by a doubling of the CO_2 concentration to be 3.9 W/m^2. As a basis for the calculations, we consider sunshades that are able to shade 1% of the inflow and estimate at which level of CO_2 concentration the warming effect then could be mitigated. Arrhenius (1896) showed that the greenhouse effect is approximately logarithmic in the CO_2 concentration. A reduction of the energy outflow by 1 percent of the inflow, i.e., 3.4 W/m^2, would then arise at a CO_2 concentration of 528 ppm. Over the last decade (not including the Corona-year) the CO_2 concentration has increased by 2.4 ppm/year. Reaching 528 ppm, producing a greenhouse effect of 3.4 W/m^2 would thus take approximately 50 years at the current speed of concentration increase.

If the sunshades were to be placed close to Earth, they would need to cover an area equal to one percent of the projected area of the Earth towards the Sun (π times the square of Earth's radius).[8] However, the distance to the Lagrange point is so large so the relevant comparison is the projected area of the Sun. Seen from Earth, it should look as if the shades cover one percent of the face of the sun. Of course, since the distance to the sun is much larger than to the sunshades, the actual area is much smaller than 1% of the Sun's.[9] To shade a share ΔS of the inflow of sunlight, the required sunshade area is equal to

$$\Delta S * (R_{dist})^2 * A_{Sun},$$

where is R_{dist} is the distance to the shades relative to the distance to the Sun and A_{Sun} is the projected area of the Sun. Using $R_{dist} = 0.0158$ and $A_{sun} = 1.50 * 10^{12}$ km^2 yields that the sunshades need to be 3.8 million km^2. This is large, seven times the area of European France or more than five times the area of Texas. Certainly, a sunshade in one piece of that size is highly impractical. Instead, the suggestion is to use shades with an individual size about as large as a football field. In the order of 500 million such shades would then need to be constructed.

[8] Of course, this would be highly impractical since the shaded area would be dark.

[9] Recall that from Earth, it looks like the Moon and the Sun are approximately as large on the sky due to the fact that the Sun is approximately 400 times further away.

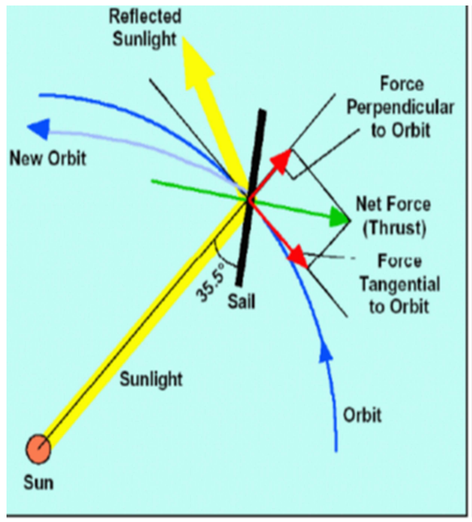

Figure 3: Solar radiation pressure used to come closer to sun.

4.3 How to get the sunshades to their final destination?

The first and foremost challenge when transporting the sunshades to their destination is to lift them up into low orbit at an altitude of about 2000 km, above the crowded space used by communication satellites. Lifting the shades to orbit will be done by reusable rockets. From this orbit, each sunshade will use the solar radiation pressure to reach their final destination, 2.36 million km away from Earth. Thus, the shades need not carry any propellant. The basic principle making this voyage possible, is that the solar radiation pressure can be used to reduce the shades speed around the Sun. The graviational force of the Sun will then dominate the centrifugal force and the shade starts to fall towards the Sun (see Figure 3). The solar radiation pressure creates a force quite weak relative to that of rockets. Nevertheless, the full journey is calculated to take no more than a few years.

As discussed above, reflectivity should be minimized when the shades are in place at their final destination, since the solar radiation pressure there should be minimized. However, a low reflectivity and the corresponding low solar radiation pressure makes the time to reach to the final destination longer. To circumvent this trade-off, the shades should be constructed with different reflectivities on its two sides. In place, the low reflectivity side would face the Sun but the other side, with high reflectivity, is used to propel the travel from Earth. While in place, the solar radiation pressure will be used to maneuver the individual sunshades, pointing them to the sun and avoiding collisions using swarm technology. Since the Largange point is unstable, the ability to maneuver is also required to keep the sunshades in place.

4.4 Cost estimation and launch plan

The main cost of this plan B in the form of sunshades is to launch the sunshades into low Earth orbit with rockets. When state agencies where the only ones to launch large rockets, the cost of lifting payload to low Earth orbit was in the order of 10 000 US$/kg. As noted above, the cost has fallen substantially thanks to reusable rocket stages and cost conscious private companies. SpaceX, the leading private space company, currently offers launches with Falcon Heavy at 90 million US$. This rocket takes a payload of 63 800 kg, implying a cost of 1400 US$/kg. To make the sunshade plan fly, costs need to fall further. It is certainly difficult to make estimates of how much further costs can fall. Here, we use a cost of 50 US$/kg as an estimate. This is about five times the current cost of the propellant (methane and liquid oxygen) and more than twice the estimate done by the CEO of SpaceX of 20 US$/kg.

As discussed above, the optimized mass per m^2 depends on the reflectivity of the sunshades. With a sufficiently low reflectivity, the optimum is achieved at an areal density of 8.8 grams/m^2 (8800 kg/km^2). Given the required area of 3.8 million km^2 and using the guesstimate 50 US$/kg the cost of lifting the sunshades to low Earth orbit is 1.7 trillion dollars. At current costs of 1400 US$/kg, the lift cost is 47 trillion dollars. Manufacturing and additional costs are arguably somewhat easier to estimate and are here set to 1.3 trillion dollars making a total of 3 trillion dollars.

The optimized mass of the sunshades is quite sensitive to the reflectivity that can be achieved. With a less optimistic assumption about the reflectivity, using current technology, the optimized mass of the sun shades is about two and a half times higher. Since the lift and manufacturing costs scale with mass, also the cost then increase by approximately a factor 2.5.

Assuming a payload capacity of 100 tons, and the need to lift 34 million tons, 330 000 launches would be required in the optimistic scenario. If this is done over a 20 year period, 46 rockets need to be launched every day.

4.4.1 Discussion

Are sunshades a reasonable plan B in case plan A against climate change fails? Certainly, this is a question that cannot be affirmatively answered without much more research. However, a few points make us believe that the idea should not prematurely be put in the drawer.

First, the estimates given above does not point to not prohibitively high costs. The optimistic calculation of 3 trillion dollars spread globally and over 20 years is of an order of magnitude that is negligible in a possible future situation where climate change threatens the sustainability of our civilization. In fact, it is small enough to be born by a large country like US or a group of countries like EU on their own. US's GDP is estimated to be 23 trillion dollars in 2021 and in EU the corresponding figure is 18 trillion (22 trillion in international purchasing power). Of this a bit over 20% is invested. Thus, both the US and the EU currently invests more every year than would be invested over a 20 year period to get the sunshades operative. In fact, even a cost an order of magnitude larger, for example due to very limited technological developments in space technology, appears manageable although probably requiring global cooperation. That the cost is manageable for a EU or the US individually is an advantage if international climate cooperation breaks down. That it is large enough to deter smaller countries may in fact be an advantage in that competing systems with different aims might be detrimental.

Second, test launches of sunshades propelled by solar radiation pressure could be done soon. The same propulsion technology is already used in practice in the crowd-funded project Lightsail 2, which in 2019 sent up a satellite in an orbit controlled by a sail propelled by solar radiation pressure. Experimental sunshades using this propulsion technology could be launched very soon at low cost.

Third, the effects on incoming sunlight of the sunshades is easily controllable. Both the total amount of sunlight deflected and its distribution over the globe can be changed basically instantaneously. Thus, the technology does in itself not induce permanent or hard to reverse effects.

Certainly, any solar geoengineering method will have side effects that must be understood (see Irvine et al., 2016, for an overview). However, sunshades in space is in principle much simpler and not as complex as e.g. stratospheric aerosol injection (SAI) (Irvine, 2016). SAI have several unknown side effects which must be carefully studied and understood before it can be implemented in full scale and will require continuous operation for centuries. Sunshades on the other hand, will stay for very long times and are in addition very difficult to attack by e.g. terrorists. Nevertheless, research on different alternatives of solar geoingeneering should proceed. History suggests that political decisions to bet on one technlogy only often leads to mistakes. A better strategy is to make final technology choices as late as possible when the developement process is not too costly.

5 Policy conclusions and suggestions for future research

Climate neutrality by 2050 is part of many countries' climate plans. Global climate neutrality by the mid of the century also appears to be a possible outcome in the international game of policy coordination. If realized, such a transition would likely lead to moderate climate damages. If the implementation is based on carbon pricing, it will steer technological change in a way that makes the transition possible without compromising growth and economic catch-up of developing countries. This should therefore be climate policy plan A.

However, due to large uncertainties, plan A may turn out to be insufficient due to stronger than expected climate sensitivity and less potential for adaptation. International policy coordination may also fail. This creates an argument for a plan B, which realistically cannot be a quick-stop of all fossil fuel use. There are theoretically valid arguments against the development of such plans. In this chapter we have characterized some, namely that they can make it more costly to achieve plan A as well as making it less robust to the possibility of multiple equilibria.

In our analysis, the costs associated with the development of plan B is costly only in the state of nature where damage sensitivities are weak while the benefits arise in the more concerning state with strong damage sensitivities. This limits the strength of the argument against a plan B. Nevertheless, our model is not quantitative and work of that sort is strongly needed. Arguably, quantitative research that demonstrates that the costs are not exceedingly large should proceed a decision on developing different potential plan B:s. However, in contrast to our stylized model, where development is an instantaneous decision, the development takes time. Therefore, we believe that experimentation on different variants of solar radiation management as well as other forms of geoengineering should not be postponed.

In this paper, we have neither discussed Carbon Capture and Storage (CCS) of CO_2 from points of emission nor Direct Air Capture (DAC) of CO_2 from the atmosphere. These methods are somewhere in between plan A and B since they, in particular CCS, are likely to be an important part of plan A. In principle, technology might advance so that the latter could form a viable plan B although costs today are prohibitive.

Finally, we want to stress that whether the sunshade plan as well as other similar ideas actually work is still highly uncertain. Sunshades may have unforeseen side effects and will not remove all negative consequences of CO_2 emissions. In particular, the acidification of the oceans resulting from increased uptake of CO_2 has known albeit quantitatively uncertain negative consequences for coral reefs and the marine life in general. A way of reducing uncertainty is to develop a portfolio of different potential plan B's. However, large degrees of uncetainty will remain. This provides a strong argument against using sunshades and other similars plans as an alternative plan A. The latter should remain a transition to carbon neutrality by the mid of the current century.

6 References

Acemoglu, Daron and Will Rafey, (2019), "Mirage on the Horizon: Geoengineering and Carbon Taxation Without Commitment", Mimeo, MIT.

Aghion, Philippe Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John Van Reenen, (2016), "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry", *Journal of Political Economy*, 124:1.

Arrhenius, Svante, (1896), "On the influence of carbonic acid in the air upon the temperature of the ground", *Philosophical Magazine and Journal of Science.* 41, pp. 237-76.

Fuglesang, Christer and Maria Garcia de Herreros Miciano, (2021), "Realistic sunshade system at L1 for global temperature control", *Acta Astronautica,* 186, pp 269-279.

Fuss, S., W. F. Lamb, M. W. Callaghan, J. Hilaire, F. Creutzig, T. Amann, T. Beringer, W. de Oliveira Garcia, J. Hartmann, T. Khanna, G. Luderer, G. F. Nemet, J. Rogelj, P. Smith, J. L. Vicente Vicente, J. Wilcox, M. del Mar Zamora Dominguez, and J. C. Minx, (2018) "Negative emissions—Part 2: Costs, potentials and side effects", *Environmental Research Letters*, 13.

Hassler, John, Per Krusell and Conny Olovsson, (2018), "The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate", *Annual Review of Economics*, 10, pp. 189—205.

Hassler, John, Per Krusell and Conny Olovsson, (2021), "Directed technical change as a response to natural-resource scarcity", *Journal of Political Economy,* forthcoming.

Howard, P., and T. Sterner (2017). "Few and not so far between: A metaanalysis of climate damage estimates". *Environmental and Resource Economics*, 68: 197–225.

IMF, (2020), *World Economic Outlook*, Chapter 3, October 2020.

IPCC, (2007), "Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change", Eds. Bert Metz, Ogunlade Davidson, Peter Bosch, Rutu Dave, and Leo Meyer.

IPCC, (2013), "Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the

Intergovernmental Panel on Climate Change", eds. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, (2021), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate 29 Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou, (eds.). Cambridge University Press. In Press.

Irvine, P., B. Kravitz, M. G. Lawrence, and H. Muri, (2016), "An overview of the Earth system science of solar geoengineering", *WIREs Climate Change*, 7:815–833.

KVA, (2018), "Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018", The Committee for the Prize in Economic Sciences in Memory of Alfred, available at https://www.nobelprize.org/uploads/2018/10/advanced-economicsciencesprize2018.pdf

Lapenis, Andrei G., (1998), "Arrhenius and the Intergovernmental Panel on Climate Change", *EOS,* Volume 79, Issue23 9 June 1998, pp 271-271.

Matthews, H. Damon, Nathan P. Gillet, Peter A. Stott and Kirsten Zickfeld (2009), "The proportionality of global warming to cumulative carbon emissions", *Nature*, Vol 459 June, pp 829-33.

Nordhaus W, Moffat A., (2017). A survey of global impacts of climate change: replications, survey methods and a

statistical analysis. NBER Work. Pap. 23646.

Pindyck, Robert S., (1991), "Irreversibility, Uncertainty, and Investment",*Journal of Economic Literature,* Vol. 29, pp. 1110-1148.

Pindyck, Robert S., (2013), "Climate change policy: What do the models tell us?", *Journal of Economic Literature* 51(3): 860–72.

Rudik, Ivan (2020), "Optimal Climate Policy When Damages are Unknown", *American Economic Journal: Economic Policy*, 12(2): 340–373.

Weitzman, Martin L., (2009), "On Modeling and Interpreting the Economics of Catastrophic Climate Change", *The Review of Economics and Statistics,* 91 (1): 1–19.