Journal of Environmental Economics and Management 84 (2017) 1–17

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Are there reasons against open-ended research into solar radiation management? A model of intergenerational decision-making under uncertainty

Martin F. Quaas [a,*], Johannes Quaas [b], Wilfried Rickels [c], Olivier Boucher [d]

[a] *Department of Economics, University of Kiel, Germany*
[b] *Institute for Meteorology, Universität Leipzig, Germany*
[c] *Kiel Institute for the World Economy, Kiel, Germany*
[d] *Laboratoire de Météorologie Dynamique, Université P. et M. Curie/CNRS, Paris, France*

ARTICLE INFO

Article history:
Received 2 May 2016
Available online 6 February 2017

JEL classification:
Q54
O32
D80

Keywords:
Climate economics
Climate engineering
Geoengineering
Solar radiation management
Learning
Hyperbolic discounting
Intergenerational decision-making
Precautionary effort

ABSTRACT

Solar radiation management (SRM) has been proposed as a means of last resort against dangerous climate change. We propose a stylized model of intergenerational decision making on SRM research, greenhouse-gas abatement and SRM deployment, under uncertainties about (a) the extent of future climate damage and (b) effectiveness and potential harmful side-effects of SRM. Open-ended research may reveal either that SRM effectively reduces climate damage, or that it would cause more harm than benefits. We find that SRM research increases the likelihood of deployment ("slippery slope"), and derive conditions that it decreases abatement effort in expectation ("moral hazard"). Neither of these provides a rationale against SRM research, though. The rational decision is to perform SRM research, unless (i) discounting is hyperbolic and (ii) the absolute prudence of expected climate damage is smaller than absolute risk aversion. These results generalize to the case where SRM research also provides information on climate sensitivity.

Introduction

Anthropogenic climate change is potentially very harmful for humankind and yet, projections of future climate change are still highly uncertain (IPCC, 2013). So far, the global society has been unable to effectively reduce carbon dioxide emissions despite some effort (e.g. UNFCCC, 1992, 1997), and ambitious objectives (UNFCCC, 2015). Solar radiation management (SRM), for example implemented by injecting sulfur into the stratosphere or by marine cloud brightening, has been proposed as an alternative high-leverage climate engineering option to keep global warming within tolerable limits (Latham, 1990; Crutzen, 2006). Indeed, the fact that climate predictions are so uncertain, and that subsequently it might become urgent to combat catastrophic climate change, has been an important argument in favor of SRM research (Crutzen, 2006; Weitzman, 2009, 2011; Keith et al., 2010; Nordhaus, 2011; Long et al., 2011; National Research Council, 2015). SRM, in that view, could be an option of last resort against catastrophic climate change. However, considering known and unknown

* Corresponding author.
E-mail address: quaas@economics.uni-kiel.de (M.F. Quaas).

http://dx.doi.org/10.1016/j.jeem.2017.02.002
0095-0696/

side-effects, it is possible that SRM as a potential "cure" of climate change could be worse than the "disease" (Schneider, 1996; Corner and Pidgeon, 2010; Robock, 2008). Against this background, one may argue that more research is needed to learn about the effectiveness and potential harmful side-effects of the various proposed SRM measures (Royal Society, 2009; Keith et al., 2010; Long et al., 2011; National Research Council, 2015). As a matter of fact, this is the question on climate engineering that requires a societal decision in the near future.

Indeed, a lively debate has emerged whether or not SRM research is desirable (Cicerone, 2006; Victor, 2008; Goodell, 2010; Robock et al., 2013; Klepper and Rickels, 2014). Probably the most prominent argument against SRM research is that the knowledge about its prospects to combat the consequences of climate change may weaken the political will to mitigate greenhouse gas emissions, causing therefore some kind of "moral hazard" behavior (Jamieson, 1996; Keith, 2000; Robock, 2008; Royal Society, 2009; Morrow et al., 2009; Gardiner, 2010; Rayner et al., 2013; Morrow, 2014; Barrett, 2014). As a consequence SRM research might push society onto a "slippery slope", creating an internal dynamic which leads to deployment, irrespective of reasons against deployment (Jamieson, 1996). Against this background, a prominent proposal in the governance debate about climate engineering SRM research is to establish a (temporal) moratorium which prohibits certain research activities, in particular those involving field testing of SRM technologies (Cicerone, 2006; Davis, 2009; Convention on Biological Diversity (CBD), 2010; Hale and Dilling, 2012; Robock, 2012; Parson and Keith, 2013; Züern and Schäfer, 2013; Schäfer et al., 2013).

In this paper, we develop a stylized theory of intergenerational decision making to scrutinize arguments for or against researching into SRM in the near future. In contrast to the literature that often restricts SRM research to the *development* of SRM technologies, we model SRM research as open-ended, that means research may reveal either that SRM is a sufficiently safe technology or that it has harmful side effects. Given that several proposals for SRM technologies are already available, this acknowledges the fact that the main scientific uncertainty is about the effectiveness and potential harmful side effects of such technologies.

We consider three generations, where the first decides on whether or not to do research on the safety or harmfulness of SRM (e.g. by means of field testing). The second generation decides on both abatement efforts and whether or not to do SRM research. The third is exposed to the consequences of climate change and decides on the deployment of the SRM technology. The decisions of both the second and the third generations will depend on the information on SRM. Decisions of the first generation takes place under the two risks of (a) an unknown extent of future climate change and (b) potential harmful side-effects of SRM.[1] The latter uncertainty could be reduced by researching into SRM. In such a situation, the value of information may be negative when the decision-maker's preferences are dynamically inconsistent (Brocas and Carrillo, 2000). This means, the optimal decision by the present generation can actually be against SRM research. In this paper we characterize the conditions for such an outcome, with regard to social time and risk preferences, greenhouse-gas mitigation technology, uncertain dynamics of climate change, and uncertainty about potential harmfulness of SRM technology. In addition, we extend the analysis by allowing for additional learning on the true value of climate sensitivity from SRM field tests.

SRM can be thought of as an insurance against catastrophic climate change because it is expected to achieve a relatively cheap and fast (i.e. within a matter of months) reduction of global temperatures (Matthews and Caldeira, 2007; Keith et al., 2010). However, there are still large uncertainties about the effectiveness of SRM measures (Ferraro et al., 2014; Klepper and Rickels, 2014). Climate models that explicitly include specific SRM measures suggest that the desired reduction in solar forcing might not be achieved that easily (Kuebbeler et al., 2012 study stratospheric sulfur injection, Korhonen et al. (2010), Stuart et al. (2013); and Alterskjaer and Kristjánsson (2013) model marine stratus cloud seeding, and Storelvmo et al. (2013) look at cirrus cloud modification). Furthermore, SRM affects the global climate system differently than reductions in the atmospheric greenhouse gas concentration. SRM can thus change precipitation patterns (Allen and Ingram, 2002; Ricke et al., 2012), and is likely to have quite different welfare effects around the globe (Royal Society, 2009; Klepper and Rickels, 2014; Quaas et al., 2016).

While these considerations and contributions have enriched the debate on the specific research challenge associated with SRM, the underlying trade-offs and implications of the research decision have rarely been investigated in a theoretical decision making framework. As one of the few exceptions, Goeschl et al. (2013) consider a two-generation model where the first generation decides on whether or not to make a SRM technology available to the second generation, which is assumed to be biased either in favor or to the disadvantage of adopting the SRM technology, compared to the benchmark of identical preferences. Because of the different preferences of the two generations, this gives rise to time inconsistent decision-making and thus a strategic game between the two generations, where the first may choose not to "arm the future" with the SRM option if (a) the second is strongly biased in favor of SRM deployment and (b) the costs for research and developing the SRM technology are sufficiently large. Our approach differs from, and extends, Goeschl et al. (2013) in several respects. First, we assume identical preferences for all generations, i.e. the third generation will deploy the SRM technology if and only if the first generation would do this in the same situation. Second, as mentioned above, we model SRM research as open-ended, i.e. the result may also be that SRM is harmful while allowing for the case that the SRM technology is available even without research, but then without information about its potential harmfulness. Goeschl et al. (2013), by contrast, assume that the

[1] Moreno-Cruz and Keith (2013) and Emmerling and Tavoni (2017) do integrated assessment modeling analyses of abatement and SRM deployment under these two uncertainties, but they do not study SRM research.

first generation directly decides on whether or not to make the SRM technology available to the future. Third, we find that the specific assumptions on damage and abatement cost functions can make a big difference for the outcome. Thus, we consider general functional forms, while Goeschl et al. (2013) assume quadratic damage and abatement cost functions.

Finally, one may find that the costs of research and development of SRM technologies are small, compared to large potential benefits of SRM. Our setup allows us to derive conditions where the social willingness to pay for SRM research is negative. In extension to Goeschl et al. (2013), we show that the optimal choice for the first generation may be to abstain from research on SRM even if that research does not cause any direct costs. We find, however, that necessary conditions for such an outcome in our framework are (i) discounting is hyperbolic and (ii) the absolute prudence of expected climate damage is smaller than absolute risk aversion. Our analysis thus highlights the important role of hyperbolic time preferences, and more generally time-inconsistent decision making, for understanding the pros and cons of open-ended SRM research, which is not yet adequately reflected in the literature. Moreover, it highlights the interaction of multiple risks in intergenerational decision-making. A prudent second generation would strongly abate emissions in case research reveals that SRM is not a way out of severe climate damage. If the degree of prudence is sufficiently high, SRM research could even reduce expected climate damage from the first generation's point of view under the two risks of (a) an unknown extent of future climate damage and (b) potential harmful side-effects of SRM. If either of these two conditions – time-inconsistent preferences and a sufficiently low degree of prudence – is violated, the best decision is to research in SRM, and to endow future generation with information on its safety or harmfulness. These results generalize to the case where SRM research provides information on climate sensitivity as well.

The next section presents the stylized model of intergenerational decision-making and learning about SRM. In Section 3 we derive our main results. Section 4 extends the analysis to include learning about climate sensitivity. The final section discusses the results and concludes.

Stylized model with three generations of decision makers

We consider three subsequent generations of decision makers. The relevant decision of the first generation is on whether or not to conduct in-depth, open-ended research on SRM, for example by means of large-scale field experiments and climate modeling.

The relevant decision of the second generation is on the amount of climate change mitigation, i.e. the quantity A of greenhouse gas emission abatement. We directly measure this quantity in monetary units, i.e. in terms of dollars or euros spent on reducing greenhouse gas emissions. While all results translate in a straightforward way to a model formulated with abatement in terms of mitigated tons of carbon, the present formulation facilitates notation and interpretation. The second generation also decides on whether or not to conduct in-depth research on SRM, in case the first generation has decided against SRM research.

The relevant decision of third generation is on whether or not to deploy SRM. Climate damage materializes just before the third generation enters the scene. Climate damage is uncertain and can turn out to be high, $d \gg 0$, or low, in which case it is normalized to zero. The probability $p(A)$ of high damage decreases with the mitigation effort A, that is $p'(A) = \partial p(A)/\partial A < 0$. We further assume $p''(A) > 0$ to assure that first-order conditions characterize a local minimum.

Due to data scarcity, it is very difficult to come up with a reliable estimate of damage functions (Pindyck, 2013), and the various integrated assessment models use quite different functional forms. Because of this uncertainty, we keep the model general with respect to the functional form of $p(A)$ for most of the analysis. For some results, we make use of the following rather flexible specification, which we formulate as

Assumption 1. The probability of severe climate change as a function of mitigation expenditures A is given by $p(A) = \left(1 - \frac{1-\beta}{\alpha} A\right)^{\frac{1}{1-\beta}}$, with $\alpha, \beta > 0$.

Assumption 1 implies $p(0) = 1$, which somewhat simplifies the expressions below, but comes without loss of generality (essentially it is just a re-normalization of d for $A=0$).[2] The assumption $p'(A) < 0$ and $p''(A) > 0$ are fulfilled for $A < \alpha/(1-\beta)$ if $\beta < 1$ and for any value of $A > 0$ if $\beta > 1$. The assumption $\beta < 1$ thus means that the probability of severe climate change can be driven to zero with finite abatement effort. The parameter β also controls the curvature properties of the $p(A)$ function; we come back to this below. Special cases of this specification are $p(A) = 1 - A/\alpha$ for $\beta = 0$, and $p(A) = \exp(-A/\alpha)$ for $\beta = 1$.

Damages are expressed in utility terms, i.e. $p(A)d$ denotes the expected utility loss from climate damage. In other words, the value of d takes into account risk aversion.[3] Since the third generation experiences climate damage, they know whether

[2] Consider the specification $\tilde{p}(A) = p_0\left(1 - \frac{1-\beta}{\alpha} A\right)^{\frac{1}{1-\beta}}$. In this case $p(0) = p_0$, which may differ from one. Expected damage for $A=0$ then is $p_0\, d$. Using the re-normalized value $\tilde{d} \equiv p_0\, d$ for the damage parameter yields the model considered in the following.

[3] This can be seen as follows. Let Y_+ denote the third generation's income (=global production output) in the case of moderate climate change and Y_- income in the case of severe climate change. Nordhaus DICE model, for example, models climate damage as a proportional reduction of global production output by a factor $\left(1 - \Psi\, T^2\right)$, where $\Psi > 0$ is a parameter and T is the global mean surface temperature above pre-industrial level (Nordhaus and Sztorc, 2013). While under moderate climate change this damage factor would be small, and hence global production output would be relatively large, the damage factor would be large under severe climate damage, and hence global output would be reduced to $Y_- < Y_+$. We assume that the third generation's risk

climate damage is low or high. In the case of low climate damage, there is no need for SRM by assumption. When climate damage is high, the third generation may consider SRM deployment. We assume that this is a decision of either a large-scale deployment or no implementation of SRM at all.

With a probability q, SRM is effective in reducing climate damage from d to a level $\underline{d}$ that is smaller than d, but still positive since there are unmitigated consequences of climate change such as ocean acidification (Williamson and Turley, 2012; Aswathy et al., 2015), $0 < \underline{d} < d$. With probability $1 - q$, SRM is not effective and also generates harmful side effects (Kravitz et al., 2013), such that damages are actually increased to a level $\overline{d} > d$.

Learning about SRM in our setting means that the decision maker knows for sure which of the two possibilities will actually materialize, i.e. whether damage will be $\overline{d}$ or $\underline{d}$ if SRM is deployed. In the framework considered here, this assumption that learning is complete is without loss of generality. All results generalize in a straightforward way to a setting where SRM research reveals which is the correct prior for the distribution of expected damages under SRM.[4]

If the third generation has no prior information about the effect of SRM, the expected damage cost of deploying SRM measures thus are $q\,\underline{d} + (1 - q)\overline{d}$. Again, the values of $\underline{d}$ and $\overline{d}$ are including the effects of risk aversion. In addition, they include the costs of SRM efforts. Finally, SRM may cause further damages in the more distant future if it is terminated for some reason (Jones et al., 2013). The risk of termination and the associated damages can also be factored in the parameters q, $\underline{d}$ and $\overline{d}$.[5]

The first or second generation can decide to conduct in-depth SRM research. This would endow the subsequent generations with the information whether SRM is safe, i.e. leading to a reduction of climate damage to $\underline{d}$, or harmful, i.e. amplifying climate damage to $\overline{d}$. The cost of such research is denoted by K, which is assumed to include direct expenditures for researchers and equipment, as well as a monetary measure of environmental damage costs of field experiments, including a risk premium as those are likely to be uncertain. The structure of decisions is illustrated in Fig. 1.

Each generation takes the subsequent generation's expected utility into account, discounted by a factor ρ, $0 < \rho < 1$. Using U^i to denote the (expected) utility generation i enjoys for itself, and W^i to denote the objective function for generation i that takes into account future generations' well-being, generation 2's objective function is $W^2 = U^2 + \rho\, U^3$, while generation 1's objective function is

$$W^1 = U^1 + \rho\left(U^2 + \delta\, U^3\right). \tag{1}$$

When deciding on SRM research, the first generation also takes the third generation's expected utility into account. We allow for hyperbolic discounting, i.e. we consider the discount factor $\rho\,\delta$ for the third generation's utility in the first generation's objective function, and allow for $\delta > \rho$. Under geometric discounting, $\delta = \rho$, the third generation's utility would be discounted using a factor ρ^2 in the first generation's objective function, which reflects that the same intergenerational discount factor is applied twice. Under hyperbolic discounting, the relative weight of utilities of the first two generations (captured by the discount factor ρ) is smaller than the relative weight of utilities of the following two generations (captured by the discount factor $\delta > \rho$).[6]

There may be good reasons to consider the case of hyperbolic social time preferences, $\rho < \delta$. Gerlagh and Liski (forthcoming) argue that considering hyperbolic time preferences reconciles discounting of far-distant climate change impacts (Stern, 2007) and realism of shorter-term decisions (à la Nordhaus). Another argument for hyperbolic discounting is the following (Heal and Millner, 2013): Consider a society of two (or more) citizens (1 and 2), who have their own preferences on intergenerational decision making, characterized by standard geometric discounting. They differ, though, in their discount factors ρ_1 and ρ_2, such that citizen i values utility levels of the three generations according to $W_i^1 = U^1 + \rho_i\, U^2 + \rho_i^2\, U^3$. The Utilitarian welfare function is the average of the W_i^1, and in the two-citizen example given is equivalent to (1) with

(*footnote continued*)
preferences are characterized by the von Neumann Morgenstern axioms, and described by an expected utility function $p(A)u(Y_-) + (1 - p(A))u(Y_+)$. Since the von Neumann Morgenstern expected utility function is defined up to linear affine transformation, we can set $u(Y_+) = 0$ without loss of generality and define $d = -u(Y_-) > -u(Y_+) = 0$ (this follows as $u(\cdot)$ is an increasing function). Given Y_+ and Y_-, a higher risk aversion would increase the difference between $u(Y_+)$ and $u(Y_-)$ and thus, in our notation, increase d.

[4] To see this, consider the setting from footnote 3, and use $\overline{Y}_-$ (respectively, $\underline{Y}_-$) to denote the third generation's income under severe climate change and safe (respectively, harmful) radiation management. There is ambiguity in the sense that there are two priors for the probability of harmful radiation management; $\overline{q}$ and $\underline{q} < \overline{q}$. Decision makers in generations 1 and 3 attach a probability q to the prior $\underline{q}$. Research on SRM reveals which one is the correct prior. Now we define $\overline{d} = -\left(\overline{q}\, u(\underline{Y}_-) + (1 - \overline{q})u(\overline{Y}_-)\right)$ and $\underline{d} = -\left(\underline{q}\, u(\underline{Y}_-) + (1 - \underline{q})u(\overline{Y}_-)\right)$ to be the expected utilities, $\overline{d}$ for the prior with high probability of harmful SRM, and $\underline{d}$ for the prior with low probability of harmful SRM.

[5] In order to factor in hyperbolic discounting by generations two and three, just as we do for generation one below, we have to take into account at least two further generations after generation three. Under hyperbolic discounting, the third generation puts a lower weight on the damages from climate change and SRM that accrue to these future generations than generation two from their point of view. In our modeling set up this means that d, $\underline{d}$, and $\overline{d}$ are the present values of damages from generation three onwards from generation two's point of view, i.e. applying the discount factor δ to damages suffered by generations four and further in the future. The corresponding damage values from generation three's point of view are *smaller* than these values, as the third generation discounts damages for their successor generations with the smaller short-run discount factor. For the purpose of the present paper, we only need to consider the damage parameters from generation two's point of view, as long as the relative ranking of damage parameters with safe SRM, no SRM and harmful SRM is the same from the points of view of generations two and three. We assume that this is the case.

[6] Note that time inconsistent decision making occurs with hyperbolic *utility* discounting, i.e. when the time preference rate increases with the planning horizon. Declining social discount rates, which are used to discount future consumption, can occur also for standard geometric utility discounting (Groom et al., 2005; Gollier, 2008; Arrow et al., 2014; Cropper et al., 2014).

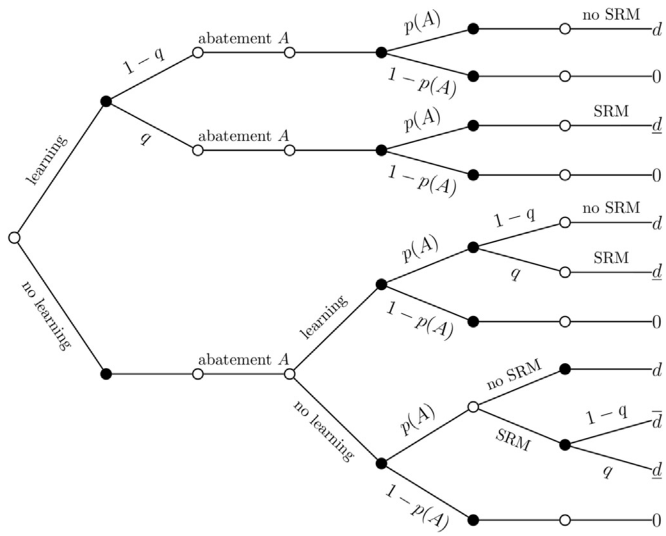

Fig. 1. Decision tree. Filled circles denote the decision points of either of the three generations; empty circles the points where uncertainty resolves.

$\rho = (\rho_1 + \rho_2)/2$ and $\delta = (\rho_1^2 + \rho_2^2)/(\rho_1 + \rho_2)$. It is straightforward to show that $\delta > \rho$ if $\rho_1 \neq \rho_2$.

In contrast to geometric discounting, hyperbolic discounting leads to time-inconsistent planning. While the first generation has some plan how the second generation should behave towards the third one, the second will deviate from this plan when it is in the actual position of decision making, because they apply the smaller discount factor ρ for the third generation's utility relative to their own, instead of δ, as the first generation does. The procedure to find the time-consistent solution, is to consider the strategic game among the generations and solve the problem as a subgame-perfect Nash equilibrium.

Analysis and results

As usual, the problem is solved backwards in time, starting with the third generation.

Third generation

For all decision nodes where climate damage turned out to be small, the third generation faces a decision problem with an obvious solution. Since they do not suffer a lot from climate damage, they opt not to deploy SRM.

With high climate damages, the third generation considers the SRM option. If both the first and the second generation have decided against SRM research (the third generation is in the lowest branch of the decision tree in Fig. 1), this is a decision under risk. The third generation compares the expected utility when deploying risky SRM, which is $q\,\underline{d} + (1-q)\overline{d}$, with the – known – utility without SRM, which is the climate damage d.

If, however, the first or second generation has provided the information about the harmfulness of SRM (the third generation is in one of the upper branches of the decision tree in Fig. 1), the third generation can use the option of SRM more effectively. Still considering the case of high climate damage, the third generation will use SRM if it is safe, such that damages are known to be only $\underline{d}$ with SRM (rather than d without). If SRM is harmful, and damage is known to be amplified by SRM to $\overline{d}$, the third generation will not use these measures and rather has to suffer the climate damage d.

Second generation

If the first generation has decided against SRM research, the second generation can still decide to do SRM research itself and provide the third generation with information about the harmfulness of SRM. We find that it will always do this if the costs of SRM research are sufficiently small.

Lemma 1. *If $K=0$, the second generation will do SRM research.*

Proof. For any abatement level A, the second generation will choose to do SRM research if $A + \rho\, p(A)\big(q\,\underline{d} + (1-q)d\big) < A + \rho\, p(A)\min\big\{d, q\,\underline{d} + (1-q)\overline{d}\big\}$. This condition is fulfilled, as $\underline{d} < d < \overline{d}$ and $0 < q < 1$.

In the following we assume that K is small enough to assure that the second generation will provide information about the harmfulness of SRM to the third generation. Most results hold similarly for the case where the second generation would not do SRM research. The main difference is that the third generation may have to decide on SRM deployment without prior information. A.4 discusses in detail the main results of our analysis for the case where generation 2 does not undertake SRM research.

The second generation decides on abatement A, taking into consideration the present value, at a discount factor ρ, of the third generation's expected damage cost. These depend on the information made available by the first generation (by SRM research). Information on SRM research done by the second generation becomes available only for generation three, though, as in Goeschl et al. (2013). Without information about the harmfulness of SRM (i.e., in the lower branch of the decision tree in Fig. 1), the second generation's decision problem is

$$\min_{A \in [0, A_{\max}]} \left\{ A + \rho\, p(A)\big(q\, \underline{d} + (1-q)d\big) \right\},$$

i.e. they minimize the present value of total cost, which is composed of the abatement expenditures A and the present value of expected climate damage cost $\rho\, p(A)\big(q\, \underline{d} + (1-q)d\big)$. This anticipates the third generation's decision on the deployment of SRM. Note that this decision problem is similar to the two-period self-protection problem recently studied in several contributions (Courbage and Rey, 2012; Eeckhoudt et al., 2012; Wang and Li, 2015): Abatement is a 'precautionary effort' that reduces the probability of a loss event occurring, i.e. provides 'self protection' in the sense of Ehrlich and Becker (1972)—except that here generation two does not protect itself, but rather the next generation, against the consequences of severe climate change.

In the optimization, the abatement expenditures have to be chosen in the feasible range $A \in [0, A_{\max}]$, with $A_{\max} \le \infty$. The first-order condition for this optimization problem is

$$1 = -\rho\, p'(A^*)\big(q\, \underline{d} + (1-q)d\big). \tag{2a}$$

Depending on the specific form of $p(A)$, corner solutions are possible. The corner solution $A^* = 0$ is a local optimum if $-\rho\, p'(0)D < 1$; the corner solution $A^* = A_{\max}$ is a local optimum if $-\rho\, p'(A_{\max})D > 1$. For the specification in Assumption 1, a corner solution $A^* = A_{\max} = \alpha/(1-\beta)$ is possible for $\beta < 1$. The optimal abatement level for an interior solution is determined by the condition that the present value of marginal expected climate damage is just equal to the marginal abatement cost (which is equal to one).

With learning about the safety of SRM in the first generation, the second and third generation know whether unmitigated damage under SRM is $\underline{d}$ or $\overline{d}$. The second generation thus can anticipate the third generation's decision. Thus, with information about the safety of SRM, the second generation's decision problem depends on whether or not SRM is safe. In the case of harmful SRM, the third generation will not make use of SRM, and the second generation's decision problem is

$$\min_{A \in [0, A_{\max}]} \left\{ A + \rho\, p(A) d \right\}$$

The first-order condition is

$$-1 = \rho\, p'(\overline{A}^*) d. \tag{2b}$$

For the case the second generation knows that SRM is safe, their decision problem is

$$\min_{A \in [0, A_{\max}]} \left\{ A + \rho\, p(A) \underline{d} \right\}$$

with first-order condition

$$-1 = \rho\, p'(\underline{A}^*) \underline{d}. \tag{2c}$$

Interior solutions are obtained for all three cases, i.e. A^*, $\underline{A}^*$, and $\overline{A}^*$ are strictly above 0 and below $A_{\max} < \infty$ if

$$-p'(0) > 1/(\rho\, \underline{d}) \quad \text{and} \quad -p'(A_{\max}) < 1/(\rho\, d). \tag{3}$$

As $\underline{d} < d$, we have the following result

Proposition 1. *Assume (3) to assure an interior solution. The optimal level of mitigation for generation two is smaller when they know that SRM is safe than in the case they do not have information about the safety of SRM, which, in turn, is smaller than in the case when they know that SRM is harmful;* $\underline{A}^* < A^* < \overline{A}^*$.

Under Assumption 1, the three abatement levels are

$$\underline{A}^* = \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) \tag{4a}$$

$$\underline{A}^{*}=\frac{\alpha}{1-\beta}\left(1-\left(\frac{\alpha}{\rho(q\,\underline{d}+(1-q)d)}\right)^{\frac{1-\beta}{\beta}}\right) \tag{4b}$$

$$\overline{A}^{*}=\frac{\alpha}{1-\beta}\left(1-\left(\frac{\alpha}{\rho\,d}\right)^{\frac{1-\beta}{\beta}}\right). \tag{4c}$$

For a given abatement level, the probability of SRM deployment would be $p(A)q$, as indicated in the decision tree (cf. Fig. 1). Yet, the second generation chooses the abatement level, and this choice depends on information about the harmfulness of SRM. As stated in Proposition 1, the probability of SRM deployment is $p(\underline{A}^{*})q$ with prior information and $p(A^{*})q < p(\underline{A}^{*})q$ without. Providing information about SRM thus increases the likelihood of SRM deployment, because of the dynamics of sequential decision-making. The following corollary to Proposition 1 can thus be interpreted as a formalization of the "slippery slope" argument.

Corollary 1 (*"Slippery slope"*). *SRM research increases the likelihood of SRM deployment.*

In our setting, the "slippery slope" occurs as a direct consequence of sequential decision making. However, the result that information about SRM creates this kind of "slippery slope" does not, by itself, provide an argument against SRM research.

We next turn to the second important argument in the discussion of SRM research, termed the "moral hazard" or "risk compensation" argument.

Proposition 2 (*"Moral hazard"*). *SRM research decreases abatement in expected terms, $q\,\underline{A}^{*}+(1-q)\overline{A}^{*}<A^{*}$, if*

$$-\frac{p'''(A)}{p''(A)}<-2\,\frac{p''(A)}{p'(A)}\quad \text{holds for all } A, \tag{5}$$

and increases abatement in expected terms, $q\,\underline{A}^{*}+(1-q)\overline{A}^{*}>A^{*}$, if (5) holds with reversed inequality for all A.

Proof. See A.1.

Unlike the "slippery slope", the "moral hazard" is not a generic result. The reason is that the outcome of open-ended SRM research, as considered here, may also be that SRM deployment would be overall detrimental. This means, SRM research may also show that SRM cannot be used as a way out of catastrophic climate change. Under this research outcome, the second generation's abatement effort is higher than in the case without information about the SRM technology. It depends on the functional form of $p(A)$ whether or not information about SRM reduces the abatement effort obtained on average over the two possible outcomes. The specific condition (5) is that the absolute prudence (Kimball, 1990) of expected future climate damage, $-p'''(A)/p''(A)$, is sufficiently small compared to the absolute risk aversion, $-p''(A)/p'(A)$.[7] Under this condition, the optimal abatement level is concave in damage, and thus the extra precautionary abatement in the case where the second generation knows that SRM is harmful is not large enough to make abatement in expected terms larger than abatement without information about SRM.[8]

Condition (5) is always fulfilled if $p(A)$ is specified as given in Assumption 1, because

$$-\frac{p''(A)}{p'(A)}=\frac{\beta}{\alpha-(1-\beta)A}>-\frac{1}{2}\,\frac{p'''(A)}{p''(A)}=\frac{\beta-\frac{1}{2}}{\alpha-(1-\beta)A}$$

for all values of A where $p(A)$ is defined. For other specifications, (5) may be violated and expected abatement with information about SRM is *higher* than under uncertainty about the harmfulness of SRM.[9]

Note also that the presence of "moral hazard" or "risk compensation", as defined here, does not provide in itself an argument against research into SRM. This is derived as a formal result in the next section.

Before turning to the first generations's decision on SRM research, we shall briefly study under which conditions a "moral hazard", as characterized in Proposition 2, induces a welfare loss for generation one. For studying this in more detail, we impose Assumption 1. Under this assumption, the abatement levels chosen by generation two are given by (4). The same expressions, but with ρ replaced by δ, specify the abatement levels that the first generation would prefer. Obviously, a "moral hazard" implies a welfare loss for the fist generation if and only if $\delta \neq \rho$. Note that only the curvature properties of $p(A)$ determine whether or not "moral hazard" occurs, preference parameters (in particular discount factors) play no role (cf.

[7] Courbage and Rey (2012, Proposition 1) show that the introduction of a background risk in the second period increases the level of precaution if and only if the decision-maker is prudent in the second period. In their model, the background risk affects both the good and the bad outcome. Here, the "background risk" about the harmfulness of SRM matters only in the bad state when climate change turns out to be damaging. Thus, Proposition 2 provides a new result for the economics of self-protection in a two-period setting.

[8] Note that $p(A)$ includes information on climate dynamics, i.e. how a certain concentration of greenhouse gases affects the probability of severe climate change, and on abatement technologies, i.e. how much abated emissions an extra dollar of abatement measures can buy.

[9] For example a specification $p(A)=1-A^{\beta}$, $\beta\in(0,1)$, would imply $-\frac{1}{2}\frac{p'''(A)}{p''(A)}=\frac{1-\beta/2}{A}>-\frac{p''(A)}{p'(A)}=\frac{1-\beta}{A}$.

Proposition 2). This demonstrates that a "moral hazard" outcome, defined as $q\,\underline{A}^* + (1-q)\overline{A}^* < A^*$, has no direct implication for (Pareto-)efficiency of abatement effort. If $\delta > \rho$, it may be the case that the second generation reduces the abatement effort more strongly, when provided with information on SRM, than the first generation would. Under Assumption 1 this is the case if and only if $\beta < 1$, as we show in A.1. Thus, we may conclude that the second generation reduces abatement effort to an "unduly" large extent, from the first generation's point of view, if and only if $\beta < 1$.

First generation

The first generation has to decide on whether or not to conduct research on the safety and effectiveness of SRM. To decide this, they compare the present value of mitigation expenditures, chosen by generation two, and of climate damages with or without SRM, suffered by generation 3 for the alternative decisions of "learning", indicated by a subscript l in the following, or "no learning", indicated by a subscript n in the following.

Without SRM research by the first generation, this present value is

$$W_n = \rho\Big(A^* + \delta\, p(A^*)\big(q\,\underline{d} + (1-q)d\big)\Big), \tag{6}$$

where A^* is the mitigation effort chosen by generation two and $p(A^*)\big(q\,\underline{d} + (1-q)d\big)$ is the third generation's expected damage. The first generation has no direct costs in this case, as no research is conducted. The second generation's utility is discounted by the factor ρ and the third generation's expected utility is discounted relative to the second generation's abatement cost by a factor $\delta \geq \rho$.

The first generation's present value of cost is composed of a few more terms when it chooses to do SRM research. The anticipated decisions by the second and third generations depend on the research outcome. They are different if SRM turns out to be safe, which has a probability q and results in a mitigation effort $\underline{A}^*$, than if it turns out to be harmful, which has the probability $1-q$ and results in $\overline{A}^*$. In addition, generation one has to bear the costs $K \geq 0$ of research. The first generation's present value of expected cost with SRM research thus is

$$W_l = K + \rho\, q\Big(\underline{A}^* + \delta\, p(\underline{A}^*)\underline{d}\Big) + \rho(1-q)\Big(\overline{A}^* + \delta\, p(\overline{A}^*)d\Big) \tag{7}$$

The first generation's willingness to pay (WTP) for SRM research is determined as follows

$$\begin{aligned}\frac{\mathrm{WTP}}{\rho} &= A^* - \Big(q\,\underline{A}^* + (1-q)\overline{A}^*\Big)\\ &\quad + \delta\Big(p(A^*)\big(q\,\underline{d} + (1-q)d\big) - \Big(q\,p(\underline{A}^*)\underline{d} + (1-q)p(\overline{A}^*)d\Big)\Big).\end{aligned} \tag{8}$$

The first generation will rationally choose to undertake SRM research if and only if WTP>K. Only with positive costs of research, $K > 0$, the discount factor from the present to the next generation, ρ, plays a role in the decision whether or not to do SRM research. Otherwise, it is only implicitly included as it affects the levels of A^*, $\underline{A}^*$, and $\overline{A}^*$. In the decision of whether or not to do SRM research, the direct costs K of research are divided by ρ. As to be expected, the costs K tend to favor a decision against SRM research. This becomes an important argument especially when including the possible damages of potential side-effects of a large-scale field experiment in K. And even more, the absolute effect of learning costs is amplified by discounting, $\rho < 1$.

Yet, in our framework (large) positive costs of research are not a necessary condition for a decision against SRM research. In the following we are particularly interested in the question if there are situations where the first generation will rationally decide against SRM research even in the case $K=0$, i.e. where the willingness to pay for SRM research is negative.

If Condition (5) for "moral hazard" holds, such that $A^* > q\,\underline{A}^* + (1-q)\underline{A}^*$, the willingness to pay for SRM research can be negative only if the last term in brackets in equation (8) is negative, i.e. if expected climate damage at A^* is smaller than expected climate damage when abatement is adapted to information about SRM. The following result states conditions under which this may be the case.

Theorem 1. *Impose (3) to assure an interior solution, and assume that "moral hazard" prevails, i.e. Condition (5) holds. The first generation has a negative willingness to pay for SRM research only if two necessary conditions are fulfilled*:

$$\delta > \rho \tag{9a}$$

$$-\frac{p'''(A)}{p''(A)} < -\frac{p''(A)}{p'(A)} \tag{9b}$$

Proof. See A.2.

The first necessary condition for a decision against SRM research is hyperbolic discounting. In line with previous findings (Brocas and Carrillo, 2000), the first generation would choose to provide costless information with geometric time preferences and the resulting time-consistent decision making. This demonstrates that time-inconsistent decision making is an important assumption for constructing a rationale against SRM research.

The second condition (9b) states that the absolute prudence of expected climate damage, $-p'''(A)/p''(A)$, must be smaller than absolute risk aversion, $-p''(A)/p'(A)$. Under this condition, expected climate damage $p(\hat{A}(D))D$, with the second generation's precautionary abatement effort $\hat{A}(D)$ adjusted to the damage in case of severe climate change, is convex in the damage parameter D. Thus, under condition (9b), the expected damage under uncertainty harmfulness of SRM is larger than expected damage with abatement adjusted to the information generated by SRM research. For this to hold, precautionary abatement must not strongly increase with climate damage, as stated in condition (9b). This is because otherwise the probability of climate damage would be greatly reduced under harmful SRM, such that expected damage would be relatively small when SRM is known to be harmful, and vice versa when SRM turns out to be safe. Also this result adds to the literature on precautionary efforts in a two-period setting (Courbage and Rey, 2012; Eeckhoudt et al., 2012; Wang and Li, 2015). Note that the necessary condition (9b) for a negative willingness to pay for SRM research is more restrictive than condition (5) for "moral hazard" to hold. This shows that the "moral hazard" itself does not provide a reason against SRM research. For specification 1, we have seen that condition (5) is always fulfilled. Condition (9b) is only fulfilled, however, if $\beta < 1$.

Prudence here is a property of the value function—expected climate damage costs depending on the present generation's abatement effort—, which crucially depends on the dynamics of the climate system and the abatement technology. Condition (9b) in particular requires that the marginal effect of abatement effort on reducing the probability of severe climate damage is not too concave. For the specification 1, this means that abatement technology must be effective enough that with sufficient abatement effort the probability of severe climate damage can be reduced to zero. Such an effective abatement technology can be imagined, for example, under learning-by-doing in abatement (Andreoni and Levinson, 2001), or in a setting with directed technical change (Acemoglu et al., 2012): If research and development efforts are directed towards the development of clean technologies ("green growth"), further R&D on clean technologies becomes more attractive, which tends to increase the effect of each dollar spend on mitigation.

We would like to emphasize that the reverse of (9b) is a sufficient condition for a positive willingness to pay for SRM research. Thus, an iso-elastic specification $p(A) = 1 - A^{\beta}$, $\beta \in (0, 1)$, would lead to the unambiguous conclusion that the first generation will choose to do costless SRM research, irrespective of the nature of time preferences.

To investigate further the conditions under which the first generation decides against SRM research, we consider the particular functional form for $p(A)$ as specified in Assumption 1. For part (b) of the following proposition, we moreover consider the limiting case $\beta \to 0$. In that case, the second generation will choose the corner solutions of either no abatement $A^* = 0$ or complete mitigation of severe climate change, $A^* = \alpha$.

Proposition 3. *(a) Under Assumption 1 and* Condition (3) *to guarantee an interior solution, the first generation has a negative willingness to pay for SRM research only if* $\beta < 1$ *and* $(1-\beta)\delta > \rho$.*(b) Under Assumption 1, in the limit* $\beta \to 0$*, the first generation has a negative willingness to pay for SRM research if* $\rho\, \underline{d}/\alpha < 1 < \rho(q\, \underline{d} + (1-q)d)/\alpha$ *and* $\delta\, \underline{d}/\alpha > 1$.

Proof. See A.3.

The first condition in Proposition 3, $\beta < 1$ is exactly the condition such that generation two reduces the abatement effort more strongly than generation one would like. Thus, an "unduly" strong reduction of abatement effort is a necessary condition for a negative willingness to pay for SRM research under the conditions of Proposition 3.

To get some idea about magnitudes, assume that the rate at which a generation discounts the next generation's utility is 2% per year, in line with Nordhaus and Sztorc (2013). The corresponding discount factor over a generation of 30 years is $\rho = 1.02^{-30} = 0.552$. For the long-term discount rate assume the figure from Stern (2007) of 0.1% per year, such that $\delta = 1.001^{-30} = 0.970$. The condition of Proposition 3(a) is thus fulfilled if $\beta < \delta/\rho - 1 = 0.75$.

We next turn to the special case of a linear specification of $p(A)$, i.e. Assumption 1 with $\beta = 0$, as in Proposition 3(b). The advantage of the corner solutions implied by the linear specification is that the interpretation becomes particularly clear. The first condition in Proposition 3(b), $\rho\, \underline{d}/\alpha < 1 < \rho(q\, \underline{d} + (1-q)d)/\alpha$, states that the second generation will choose full abatement if it does not know whether the third will deploy SRM, but that they will choose no mitigation at all if they anticipate SRM deployment by the third generation. The second condition, $\delta\, \underline{d}/\alpha > 1$, states that the first generation would prefer (full) abatement on behalf of the second generation, such that there will be no need for the third generation to deploy SRM.

For this example, consider the Utilitarian welfare function with two citizens who have different geometric discount factors ρ_1 and ρ_2. We have shown above $\rho = (\rho_1 + \rho_2)/2$ and $\delta = (\rho_1^2 + \rho_2^2)/(\rho_1 + \rho_2)$. The conditions $\rho\, \underline{d}/\alpha < 1$ and $\delta\, \underline{d}/\alpha > 1$ from Proposition 3(b) imply that the citizens must strongly disagree about the abatement in case that SRM is safe: One prefers the corner solution of no abatement, the other one the corner solution of full mitigation.[10] The second condition, $\rho(q\, \underline{d} + (1-q)d)/\alpha > 1$ states that both individuals are sufficiently concerned about the possibility of severe climate change, $(1-q)d$, that they would do full abatement in the absence of information about SRM. As their individual preferences are time consistent, both would do SRM research if they could implement their preferred abatement level. Yet, in the actual decision problem, the Utilitarian welfare function prefers no SRM research in order to commit the second generation to full mitigation of climate change, according to the preference of the more patient citizen, whose preference dominates for long-term decision making.

[10] Proof: Assume to the contrary that both prefer no abatement $\rho_1\, \underline{d}/\alpha < 1$ and $\rho_2\, \underline{d}/\alpha < 1$. This implies the first condition $\rho\, \underline{d}/\alpha < 1$, and also $\left(\rho_j^2/(\rho_1 + \rho_2)\right)\left(\underline{d}/\alpha\right) < \rho_j/(\rho_1 + \rho_2)$ for j=1,2. Adding both conditions leads to a contradiction to the last inequality in Proposition 3(b), $\delta\, \underline{d}/\alpha > 1$.

SRM research and climate sensitivity

Climate damage is a function of both the radiative forcing induced by changes in carbon dioxide and other greenhouse gases and the climate sensitivity, typically measured as the change in global mean surface temperature for a doubling of the CO_2 concentration. The likely range of climate sensitivity has been quantified at 1.5–4.5 °C (IPCC, 2013). This is a large uncertainty: For reaching a target maximum warming of +2 °C over pre-industrial temperatures (UNFCCC, 2015), the lower bound would require only moderate mitigation efforts, while the upper bound would call for immediate and far-reaching action. Thus, a better scientific understanding of the impact of emitted aerosols on the Earth's radiation budget would allow for a better quantification of climate sensitivity and subsequently of climate impact damage.

Climate sensitivity could in principle be inferred from the observed warming since pre-industrial times. Unfortunately, this is hampered by the lack of quantitative understanding of the radiative forcing due to anthropogenic aerosol emissions (Schwartz, 2012).[11] SRM schemes, in particular the propositions to exploit marine cloud brightening or to alter the properties of thin cirrus clouds, intend to make use of the pathways of aerosol forcing that are most uncertain. Learning about SRM thus implies also a better scientific understanding about inadvertent climate forcing via anthropogenic aerosol emissions, and subsequently, an improved quantification of climate sensitivity, which, in turn, implies a better quantification of the shape of $p(A)$, i.e. how abatement of greenhouse gas emission could affect the probability of severe damage.

In the following we extend our model to analyze the additional incentives to perform SRM research that arise if learning about SRM also reveals information about climate sensitivity. We assume that climate sensitivity (CS) can be lower or higher than anticipated by the first generation. This uncertainty about climate sensitivity translates into ambiguity about the probability of high damage. Depending on climate sensitivity, this probability is $\underline{s}\, p(A)$ or $\overline{s}\, p(A)$, with $\underline{s} < 1 < \overline{s}$. We further assume that the second generation is ambiguity-neutral. An extension to more general ambiguity preferences (e.g. Klibanoff et al., 2005; Traeger, 2009) is left for future research. With these assumptions, four learning outcomes are possible: SRM can be safe or harmful, and climate sensitivity may be low or high. Some of these outcomes can be correlated. We use notation for probabilities as follows:

	CS	
damage with SRM	$\underline{s}$	$\overline{s}$
$\underline{d}$	$\underline{\underline{q}}$	$\underline{q}$
$\overline{d}$	$\overline{q}$	$\overline{\overline{q}}$

We further define the optimal abatement levels and derive the resulting expressions for the Assumption 1 as follows

$$\underline{\underline{A}}^{**} = \arg\min_A \left\{ A + \rho\, \underline{s}\, p(A) \underline{d} \right\} \overset{\text{Ass. 1}}{=} \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \underline{s}\, \underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) \tag{10a}$$

$$\underline{A}^{**} = \arg\min_A \left\{ A + \rho\, \overline{s}\, p(A) \underline{d} \right\} \overset{\text{Ass. 1}}{=} \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \overline{s}\, \underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) \tag{10b}$$

$$\overline{A}^{**} = \arg\min_A \left\{ A + \rho\, \underline{s}\, p(A) d \right\} \overset{\text{Ass. 1}}{=} \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \underline{s}\, d}\right)^{\frac{1-\beta}{\beta}}\right) \tag{10c}$$

$$\overline{\overline{A}}^{**} = \arg\min_A \left\{ A + \rho\, \overline{s}\, p(A) d \right\} \overset{\text{Ass. 1}}{=} \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \overline{s}\, d}\right)^{\frac{1-\beta}{\beta}}\right), \tag{10d}$$

with $\underline{\underline{q}} + \underline{q} + \overline{q} + \overline{\overline{q}} = 1$. In Eqs. (10c) and (10d) we have used the fact that the third generation would not use SRM if they know that SRM is harmful ($\overline{d} > d$), and rather suffer damage d (instead of using it and suffering $\overline{d}$).

To allow the comparison with the previously discussed case without learning about climate sensitivity, we assume

[11] Climate sensitivity can be inferred from the observed global warming record for a given rate of ocean heat uptake (which is comparatively well known), to the extent that the radiative forcing is quantifiable. The net radiative forcing is due to the warming effect by anthropogenic greenhouse gases (which is known with a relatively small error bar), natural forcing (which are comparatively small) and the net cooling effect by anthropogenic aerosol particles. The latter effects are poorly constrained and effectively hamper the possibility to infer climate sensitivity.

$\underline{\underline{q}} + \underline{q} = q$, and that $\underline{s}$ and $\bar{s}$ are scaled such that $\underline{\underline{q}}\,\underline{s} + \underline{q}\,\bar{s} = q$ and $\bar{q}\,\underline{s} + \bar{\bar{q}}\,\bar{s} = 1 - q$. Thus, the optimal choice without learning remains the same as in the previously analyzed case.

$$A^* = \arg\min_A \left\{ A + \rho\, p(A)\left(q\, \underline{d} + (1 - q)d\right) \right\} \tag{10e}$$

Imposing Assumption 1, we can compare A^* with expected abatement if generation 2 adjusts the abatement level to information about both SRM and climate sensitivity.

Proposition 4 (*"moral hazard" under combined SRM/CS research*). *Under Assumption 1, combined SRM/CS research decreases abatement in expected terms,* $\underline{\underline{q}}\,\underline{\underline{A}}^{**} + \underline{q}\,\underline{A}^{**} + \bar{q}\,\bar{A}^{**} + \bar{\bar{q}}\,\bar{\bar{A}}^{**} < A^{\star}$.

Proof. See A.5.

This result depends on how we assumed that learning about climate sensitivity affects the decision problem, namely that CS affects both $\underline{d}$ and d equally. In general, the influence on expected abatement depends on the correlations in outcomes between harmfulness of SRM and level of climate sensitivity. Ricke et al. (2012) investigates the potential of stratospheric sulfur injection as SRM for different climate sensitivities. They show that higher climate sensitivity increases effectiveness of SRM in controlling global mean temperature, but they also show that under high climate sensitivity SRM is "least effective in returning regional climates to their baseline states and minimizing regional rates of precipitation change" (p. 95). Consequently, one could argue that the difference between $\bar{d}$ and d is differently affected by CS.

Turning again to the first generation's decision problem. The first generation's willingness to pay for combined SRM/CS research is determined by

$$\begin{aligned} \frac{\mathrm{WTP}}{\rho} = A^* - &\left(\underline{\underline{q}}\,\underline{\underline{A}}^{**} + \underline{q}\,\underline{A}^{**} + \bar{q}\,\bar{A}^{**} + \bar{\bar{q}}\,\bar{\bar{A}}^{**} \right) + \delta\Big(p(A^*)\left(q\,\underline{d} + (1-q)d\right) \\ &- \left(\underline{\underline{q}}\,\underline{s}\, p(\underline{\underline{A}}^{**})\underline{d} + \underline{q}\,\bar{s}\, p(\underline{A}^{**})\underline{d} + \bar{q}\,\underline{s}\, p(\bar{A}^{**})d + \bar{\bar{q}}\,\delta\,\bar{s}\, p(\bar{\bar{A}}^{**})d \right)\Big) \end{aligned} \tag{11}$$

The following proposition shows how the additional information about climate sensitivity, that comes with the learning about SRM, changes the first generation's optimal decision on SRM research.

Proposition 5. *Impose (3) to assure an interior solution.*

(*a*) *With geometric discounting $\rho = \delta$, additional learning about climate sensitivity increases the willingness to pay for SRM research.*

(*b*) *Under Assumption 1, the first generation has a negative willingness to pay for SRM research with additional learning about climate sensitivity only if $\beta < 1$ and $(1 - \beta)\delta > \rho$.*

Proof. See A.6.

Proposition 5 shows that the result of Theorem 1 generalizes to the case where SRM research also provides knowledge about climate sensitivity.

Discussion and conclusion

In this paper we have investigated the "moral hazard" (mitigation efforts may be unduly reduced) and "slippery slope" (once started, solar radiation management (SRM) may become the preferred means of climate policy) arguments against conducting in-depth research on SRM. We have considered a stylized theoretical model of intergenerational decision making and learning under the two risks of (a) the extent of future climate change and (b) potential harmful side-effects of SRM. Our main focus was in the question under which conditions the first generation would optimally decide against SRM research. Within our framework, the value of information obtained by open-ended research may be negative for the present generation if two necessary conditions hold simultaneously: (i) Time preferences of the current decision-makers are time inconsistent and (ii) the absolute prudence of expected climate damage is smaller than absolute risk aversion. These results are robust against the inclusion of additional learning about climate sensitivity.

Condition (i) means that the first generation's social time preferences are characterized by hyperbolic discounting and thus time inconsistent. In formal terms, the decision setting becomes a strategic game among the generations, which makes it possible that the first generation chooses not to provide information to the third one (Brocas and Carrillo, 2000; Carrillo and Mariotti, 2000; Gollier and Treich, 2003). The reason is that under hyperbolic time preferences the absence of information about harmfulness of SRM is a commitment device that forces the second generation to increase their abatement efforts.

While individual behavior and intertemporal decision-making and preferences often are characterized by hyperbolic discounting and dynamically inconsistent preferences, it is an open question whether hyperbolic discounting is sensible in the social decision on SRM research. On the one hand, time consistency seems to be an appealing requirement for social decision making, and it is a rationality requirement often imposed on time preferences (Koopmans, 1960). On the other hand, there may also be good reasons to consider hyperbolic time preferences in social decision making. Hyperbolic time preferences may allow to reconcile long-term concerns of climate change with realistic short-term decision-making (Gerlagh and Liski, 2017). Moreover, a Utilitarian welfare function that aggregates heterogeneous intertemporal preferences of citizens exhibits hyperbolic discounting, even if the time preferences of all individuals are characterized by geometric discounting (Heal and Millner, 2013).

Condition (ii) is also necessary for a negative willingness to pay for SRM research in our framework, because the research outcome may also be that SRM is ruled out as a means to combat dangerous climate change. In this case, a sufficiently prudent second generation would increase precautionary abatement effort so strongly that climate damage, in expectation over the two risks of (a) an unknown extent of future climate damage and (b) potential harmful side-effects of SRM, would actually be reduced by SRM research. Prudence, here, is a property of the value function of the dynamic optimization problem, and thus depends not only on preferences, but primarily on the abatement technology and the climate system. There may be reasons why this condition could be fulfilled, such as increasing returns to abatement expenditures in terms of mitigated climate change. These may arise, for example, if there is learning-by-doing in abatement (Andreoni and Levinson, 2001) or in a setting of directed technical change (Acemoglu et al., 2012).

We have focused on the case where the second generation would do SRM research and provide information about safety or harmfulness of SRM to the third generation. Situations are conceivable where the second generation would no do SRM research. Most of our results generalize to this case, with two qualifications (cf. A.4). First, if the third generation is sufficiently willing to take risks to deploy SRM even in the absence of prior information, the slippery slope is not a generic outcome any more, but it may rather be that SRM research by the first generation *decreases* the likelihood of SRM deployment. This occurs in particular when damages for SRM being "safe" or "harmful" are rather similar, or when the probability of SRM being harmful is large. Second, if the third generation is risk-averse enough not to use SRM without prior information, the moral hazard necessarily occurs. This is because the (severe) climate damage faced by the third generation will then be the same when SRM is known to be harmful or when the third generation has no prior information – in either case it will not use SRM. Information about safety or harmfulness of SRM can then only reduce the second generation's abatement level. While hyperbolic discounting continues to be a necessary condition for a negative willingness to pay for SRM research in the first generation, the second condition is on risk aversion embodied in expected climate change, not on prudence. This is because with equal (severe) climate damages faced by the third generation in all cases except with information that SRM is safe, only climate risk is relevant for the first generation's decision, not the change in climate risk due to learning about SRM (cf. A.4).

A definite answer to the question whether or not to conduct SRM research is beyond the scope of this paper. Yet, we believe that our analysis helps clarifying the discussion, as it reduces the peculiar moral questions connected to SRM research to issues that economists are used to handling: abatement technologies and discounting. Our results show that the most important arguments put forward in the literature against SRM research do not, by themselves, justify a rational decision against open-ended SRM research. It is only in combination with time-inconsistent decision-making, i.e., hyperbolic preferences in the framework of this paper, that an argument against SRM can be rationalized.

Appendix A

A. Proof of Proposition 2

Let $\hat{A}(D)$ be defined by $p'(\hat{A}(D)) = -1/(\rho\, D)$. Obviously $A^* > q\,\underline{A}^* + (1-q)\overline{A}^*$ if and only if $A(D)$ is concave in D. Using the implicit function theorem,

$$p''(\hat{A}(D))\frac{\partial \hat{A}(D)}{\partial D} = \frac{1}{\rho\, D^2}$$

$$\frac{\partial \hat{A}(D)}{\partial D} = \frac{1}{\rho\, D^2\, p''(\hat{A}(D))} = \rho\frac{\left(p'(\hat{A}(D))\right)^2}{p''(\hat{A}(D))} > 0$$

$$\begin{aligned}
\frac{\partial^2 \hat{A}(D)}{\partial D^2} &= -\frac{2}{\rho\, D^3\, p''(\hat{A}(D))} - \frac{p'''(\hat{A}(D))}{\rho\, D^2\left(p''(\hat{A}(D))\right)^2}\frac{\partial \hat{A}(D)}{\partial D} \\
&= -\frac{2}{\rho\, D^3\, p''(\hat{A}(D))} - \frac{p'''(\hat{A}(D))}{\rho^2\, D^4\left(p''(\hat{A}(D))\right)^3} \\
&= -\frac{\rho^2}{\rho^3\, D^3\left(p''(\hat{A}(D))\right)^3}\left(2\left(p''(\hat{A}(D))\right)^2 + \frac{1}{\rho\, D}\, p'''(\hat{A}(D))\right) \\
&= -\frac{\rho^2}{\rho^3\, D^3\left(p''(\hat{A}(D))\right)^3}\left(2\left(p''(\hat{A}(D))\right)^2 - p'(\hat{A}(D))p'''(\hat{A}(D))\right)
\end{aligned}$$

This expression is negative, i.e. $\hat{A}(D)$ is concave in D, if and only if (5) holds.

Imposing Assumption 1, we have

$$\begin{aligned}
&A^* - \left(q\,\underline{A}^* + (1-q)\overline{A}^*\right)\\
&= \frac{\alpha}{1-\beta}\left(1-\left(\frac{\alpha}{\rho\left(q\,\underline{d}+(1-q)\overline{d}\right)}\right)^{\frac{1-\beta}{\beta}} - \left(q\left(1-\left(\frac{\alpha}{\rho\,\underline{d}}\right)^{\frac{1-\beta}{\beta}}\right)+(1-q)\left(1-\left(\frac{\alpha}{\rho\,\overline{d}}\right)^{\frac{1-\beta}{\beta}}\right)\right)\right)\\
&= -\frac{\alpha}{1-\beta}\left(\left(\frac{\alpha}{\rho\left(q\,\underline{d}+(1-q)\overline{d}\right)}\right)^{\frac{1-\beta}{\beta}} - \left(q\left(\frac{\alpha}{\rho\,\underline{d}}\right)^{\frac{1-\beta}{\beta}}+(1-q)\left(\frac{\alpha}{\rho\,\overline{d}}\right)^{\frac{1-\beta}{\beta}}\right)\right)\\
&= -\frac{\alpha^{\frac{1}{\beta}}\rho^{-\frac{1-\beta}{\beta}}}{1-\beta}\left(\left(\left(q\,\underline{d}+(1-q)\overline{d}\right)\right)^{-\frac{1-\beta}{\beta}} - \left(q\,\underline{d}^{-\frac{1-\beta}{\beta}}+(1-q)\overline{d}^{-\frac{1-\beta}{\beta}}\right)\right)
\end{aligned} \tag{12}$$

This expression is positive, as condition (5) holds under Assumption 1, and thus decreasing with ρ if $\beta < 1$ and increasing with ρ if $\beta > 1$. Thus, the reduction in expected abatement effort is smaller when applying a discount factor $\delta > \rho$ if and only if $\beta < 1$.

B. Proof of Theorem 1

We first show that with geometric discounting, $\rho = \delta$, the optimal choice for the first generation is to undertake SRM research. This is true, because for $\rho = \delta$

$$\frac{1}{\rho}(W_n - W_l) = q\left(A^* + \rho\,p(A^*)\underline{d} - \left(\underline{A}^* + \rho\,p(\underline{A}^*)\underline{d}\right)\right) + (1-q)\left(A^* + \rho\,p(A^*)\overline{d} - \left(\overline{A}^* + \rho\,p(\overline{A}^*)\overline{d}\right)\right) > 0,$$

which holds as $\underline{A}^* = \operatorname{argmin}_A\left\{A + \rho\,p(A)\underline{d}\right\}$, $\overline{A}^* = \operatorname{argmin}_A\left\{A + \rho\,p(A)\overline{d}\right\}$, and $\underline{A}^* < A < \overline{A}^*$ (cf. Proposition 1).

The second necessary condition is that damage at expected abatement is smaller than expected damage, $p(A^*)\left(q\,\underline{d} + (1-q)\overline{d}\right) < q\,p(\underline{A}^*)\underline{d} + (1-q)p(\overline{A}^*)\overline{d}$. This is always the case if $p(\hat{A}(D))D$ is convex in D.

$$\begin{aligned}
\frac{d^2}{dD^2}\left(\rho\,p(\hat{A}(D))D\right) &= \frac{d}{dD}\left(\rho\,p'(\hat{A}(D))D\,\frac{\partial\hat{A}(D)}{\partial D} + \rho\,p(\hat{A}(D))\right)\\
&= \frac{d}{dD}\left(-\frac{\partial\hat{A}(D)}{\partial D} + \rho\,p(\hat{A}(D))\right) = -\frac{\partial^2\hat{A}(D)}{\partial D^2} + \rho\,p'(\hat{A}(D))\frac{\partial\hat{A}(D)}{\partial D}\\
&= -\frac{\partial^2\hat{A}(D)}{\partial D^2} - \frac{1}{D}\frac{\partial\hat{A}(D)}{\partial D}
\end{aligned}$$

The condition can thus be written as

$$-\frac{D\,\frac{\partial^2\hat{A}(D)}{\partial D^2}}{\frac{\partial\hat{A}(D)}{\partial D}} > 1$$

$$-\frac{D\,\hat{A}''(D)}{\hat{A}'(D)} > 1$$

The reaction of abatement to damage must be strong enough. Using the above results, this condition becomes

$$\begin{aligned}
-\frac{D\,\hat{A}''(D)}{\hat{A}'(D)} &= \frac{\frac{D}{\rho\,D^3(p''(\hat{A}(D)))^3}\left(2(p''(\hat{A}(D)))^2 - p'(\hat{A}(D))p'''(\hat{A}(D))\right)}{\frac{1}{\rho\,D^2\,p''(\hat{A}(D))}}\\
&= 2 - \frac{p'(\hat{A}(D))p'''(\hat{A}(D))}{(p''(\hat{A}(D)))^2} > 1
\end{aligned}$$

Which holds if and only if condition (9b) holds.

C. Proof of Proposition 3

Using Assumption 1,

$$\underline{A}^* + \delta\, p(\underline{A}^*)\underline{d} = \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\, \underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) + \delta\left(\frac{\alpha}{\rho\, \underline{d}}\right)^{\frac{1}{\beta}} \underline{d}$$

$$= \frac{\alpha - \underline{d}\left(\frac{\alpha}{\rho\, \underline{d}}\right)^{\frac{1}{\beta}}(\rho - (1-\beta)\delta)}{1-\beta}$$

Using this and similar results for the other cases in (8), the condition for a negative willingness to pay for SRM research becomes (after canceling a couple of common terms on both sides of the inequality)

$$\left(q\, \underline{d} + (1-q)d\right)^{-\frac{1}{\beta}}\left((1-\beta)\delta - \rho\right) < \left(q\, \underline{d}^{-\frac{1}{\beta}} + (1-q)d^{-\frac{1}{\beta}}\right)\left((1-\beta)\delta - \rho\right)$$

Since $q^{-1/\beta}$ is a convex function, $\left(q\, \underline{d} + (1-q)d\right)^{-\frac{1}{\beta}} > \left(q\, \underline{d}^{-\frac{1}{\beta}} + (1-q)d^{-\frac{1}{\beta}}\right)$. Thus, doing SRM research is optimal if $(1-\beta)\delta < \rho$, while a necessary condition for a negative willingness to pay for SRM research is $(1-\beta)\delta > \rho$.

D. Main results without SRM research by the second generation

In our main analysis, we assume $K=0$, such that the second generation will do SRM research (Lemma 1). It is, however, also conceivable that the second generation decides against SRM research, for example if K is prohibitively high. Thinking about a situation with more than three generations, similar reasons as given in Theorem 1 for the first generation could also induce generation 2 to decide against SRM – if the other parameters are such that the third generation would not deploy SRM without prior information.

Without SRM research by generation 2, the third generation decides about SRM deployment without the additional information about its harmfulness or safety (the lowest branch in the decision tree in Fig. 1). With severe climate change, the third generation will deploy SRM if $q\, \underline{d} + (1-q)\overline{d} < d$ (case 1) and will not deploy SRM if $q\, \underline{d} + (1-q)\overline{d} > d$ (case 2). In the following, we re-state the main results on abatement effort by generation 2 (Proposition 1), slippery slope (Corollary 1), moral hazard (Proposition 2), and willingness to pay for SRM research by generation 1 (Theorem 1) for these two cases without SRM research by generation 2.

D.1. Case 1: generation 3 uses SRM in case of severe climate change

In this case, the present value of expected climate damage cost for the case where the first generation does not conduct SRM research becomes $\rho\, p(A)(q\, \underline{d} + (1-q)\overline{d}) < \rho\, p(A)d$. The ranking of abatement levels is as stated in Proposition 1.

The probability of SRM deployment is $p(\underline{A}^*)q$ with prior information and $p(A^*)$ without. The "slippery slope" is not a generic outcome of the model in this case. Specifically under Assumption 1, SRM research by the first generation *decreases* the likelihood of SRM deployment if $p(\underline{A}^*)q < p(A^*)$

$$\begin{aligned}
&\Leftrightarrow\quad \left(1 - \frac{1-\beta}{\alpha}\underline{A}^*\right)^{\frac{1}{1-\beta}} q < \left(1 - \frac{1-\beta}{\alpha}A^*\right)^{\frac{1}{1-\beta}}\\
&\Leftrightarrow\quad q\, \underline{d}^{-\frac{1}{\beta}} < \left(q\, \underline{d} + (1-q)\overline{d}\right)^{-\frac{1}{\beta}}\\
&\Leftrightarrow\quad \frac{\overline{d}}{\underline{d}} < \frac{1-q^{1+\beta}}{q^{\beta} - q^{1+\beta}}.
\end{aligned} \tag{13}$$

Thus, no slippery slope occurs if the ratio of damages under safe and harmful SRM, $\overline{d}/\underline{d}$, is sufficiently small. It is straightforward to show that the right-hand side of (13) monotonically decreases with q. Thus, no slippery slope occurs if the probability $1-q$ that SRM is harmful is sufficiently large.

As the ranking of abatement levels remains the same as in Proposition 1, the line of argument for a moral hazard to occur remains the same as in Proposition 2, and the line of argument for a negative willingness to pay for SRM research by the first generation remains the same as in Theorem 1. Thus, Proposition 2 and Theorem 1 hold as in the case where the second generation does SRM research.

D.2. Case 2: generation 3 does not use SRM

In this case, the present value of expected climate damage cost simplifies to $\rho\, p(A)d$. The ranking of abatement levels (Proposition 1) changes, because in the case considered here the second generation chooses the same abatement level when it knows that SRM is harmful and when it has no prior information about SRM, $A^* = \overline{A}^*$. This is because in both cases the third generation will suffer from the same level of climate damage, d. The full ranking of abatement levels, replacing Proposition 2, thus is $\underline{A}^* < A^* = \overline{A}^*$.

With no SRM research by generation 2 and thus no SRM deployment by generation 3, both the slippery slope and the moral hazard are generic results of the model. The likelihood of SRM deployment necessarily increases with SRM research by generation 1, as generation 3 will not deploy SRM without information about its harmfulness. Expected abatement also necessarily decreases with SRM research by the first generation, as $q\,\underline{A}^* + (1-q)\overline{A}^* = q\left(\underline{A}^* - A^*\right) + A^* < A^*$, which always is true as $\underline{A}^* < A^*$. Thus there necessarily is this kind of "moral hazard".

The proof of Theorem 1 shows that (9a), i.e. hyperbolic discounting, remains a necessary condition for a negative willingness to pay for SRM research for the first generation also without SRM research by generation 2. Condition (9b) of Theorem 1 changes, though. The second necessary condition for a negative willingness to pay for SRM research is that damage at expected abatement is smaller than expected damage (cf. A.2). In the case considered here, this condition is $p(A^*)d < q\,p(\underline{A}^*)\underline{d} + (1-q)p(A^*)d$, or, equivalently, $p(A^*)d < p(\underline{A}^*)\underline{d}$. This condition holds if $p(\hat{A}(D))D$ is decreasing in D, i.e. if

$$-\frac{\partial \hat{A}(D)}{\partial D} + \rho\, p\left(\hat{A}(D)\right) < 0 \Leftrightarrow \frac{p''\left(\hat{A}(D)\right)p\left(\hat{A}(D)\right)}{\left(p'\left(\hat{A}(D)\right)\right)^2} < 1 \Leftrightarrow -\frac{p''\left(\hat{A}(D)\right)}{p'\left(\hat{A}(D)\right)} < -\frac{p'\left(\hat{A}(D)\right)}{p\left(\hat{A}(D)\right)}. \tag{14}$$

This condition has a similar structure as Condition (9b) in Theorem 1, but it is a condition involving only the second derivative of $p(\cdot)$. Because $A^* = \underline{A}^*$, the comparison of damage at expected abatement and expected damage is independent of the uncertainty associated to SRM research. It only depends on climate risk, not on the change in climate risk due to learning about SRM. Thus, the condition is on risk aversion (with respect to uncertainty of climate damage), and not on prudence which determines how the choice of abatement level depends on the additional uncertainty of the SRM research outcome.

E. Proof of Proposition 4

$$\begin{aligned}
A^* &= \frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho(q\,\underline{d} + (1-q)d)}\right)^{\frac{1-\beta}{\beta}}\right) \\
&> \underline{\underline{q}}\,\frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\,\underline{s}\,\underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) + \underline{q}\,\frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\,\overline{s}\,\underline{d}}\right)^{\frac{1-\beta}{\beta}}\right) \\
&\quad + \overline{q}\,\frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\,\underline{s}\,d}\right)^{\frac{1-\beta}{\beta}}\right) + \overline{\overline{q}}\,\frac{\alpha}{1-\beta}\left(1 - \left(\frac{\alpha}{\rho\,\overline{s}\,d}\right)^{\frac{1-\beta}{\beta}}\right)
\end{aligned} \tag{15}$$

$$\frac{\left(q\,\underline{d} + (1-q)d\right)^{-\frac{1-\beta}{\beta}}}{1-\beta} < \frac{\underline{\underline{q}}\left(\underline{s}\,\underline{d}\right)^{-\frac{1-\beta}{\beta}} + \underline{q}\left(\overline{s}\,\underline{d}\right)^{-\frac{1-\beta}{\beta}} + \overline{q}\left(\underline{s}\,d\right)^{-\frac{1-\beta}{\beta}} + \overline{\overline{q}}\left(\overline{s}\,d\right)^{-\frac{1-\beta}{\beta}}}{1-\beta} \tag{16}$$

If $\overline{s} = \underline{s} = 1$, this inequality holds, as $\beta > 0$ (cf. Proposition 2). Turning to $\underline{s} < \overline{s}$ we obtain from the conditions on probabilities (i.e., $\underline{\underline{q}} + \underline{q} + \overline{q} + \overline{\overline{q}} = 1$, $\underline{\underline{q}} + \underline{q} = q$, $\underline{\underline{q}}\,\underline{s} + \underline{q}\,\overline{s} = q$ and $\overline{q}\,\underline{s} + \overline{\overline{q}}\,\overline{s} = 1 - q$)

$$\begin{aligned}
\underline{\underline{q}} &= \frac{\overline{s} - 1}{\overline{s} - \underline{s}}\,q \\
\underline{q} &= \frac{1 - \underline{s}}{\overline{s} - \underline{s}}\,q \\
\overline{q} &= \frac{\overline{s} - 1}{\overline{s} - \underline{s}}(1-q) \\
\overline{\overline{q}} &= \frac{1 - \underline{s}}{\overline{s} - \underline{s}}(1-q)
\end{aligned}$$

The condition that A^* is larger than $\underline{\underline{q}}\,\underline{\underline{A}}^{**} + \underline{q}\,\underline{A}^{**} + \overline{q}\,\overline{A}^{**} + \overline{\overline{q}}\,\overline{\overline{A}}^{**}$ becomes

$$\frac{\left(q\,\underline{d}+(1-q)d\right)^{-\frac{1-\beta}{\beta}}}{1-\beta}<\frac{\left(\underline{\underline{q}}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\underline{q}\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)\underline{d}^{-\frac{1-\beta}{\beta}}+\left(\bar{q}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\bar{\bar{q}}\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)d^{-\frac{1-\beta}{\beta}}}{1-\beta}$$

$$=\frac{\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,q\,\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,q\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)\underline{d}^{-\frac{1-\beta}{\beta}}+\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}(1-q)\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}(1-q)\bar{s}^{-\frac{1-\beta}{\beta}}\right)d^{-\frac{1-\beta}{\beta}}}{1-\beta}$$

$$=\frac{\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)q\,\underline{d}^{-\frac{1-\beta}{\beta}}+\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)(1-q)d^{-\frac{1-\beta}{\beta}}}{1-\beta}$$

$$=\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,\bar{s}^{-\frac{1-\beta}{\beta}}\right)\frac{q\,\underline{d}^{-\frac{1-\beta}{\beta}}+(1-q)d^{-\frac{1-\beta}{\beta}}}{1-\beta}$$

For the first term in the last line the following holds by Jensen's inequality

$$\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,\underline{s}^{-\frac{1-\beta}{\beta}}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,\bar{s}^{-\frac{1-\beta}{\beta}}\gtreqless\left(\frac{\bar{s}-1}{\bar{s}-\underline{s}}\,\underline{s}+\frac{1-\underline{s}}{\bar{s}-\underline{s}}\,\bar{s}\right)^{-\frac{1-\beta}{\beta}}=1\quad\text{for}\quad\beta\lesseqgtr 1,$$

as $\frac{1-\underline{s}}{\bar{s}-\underline{s}}=1-\frac{\bar{s}-1}{\bar{s}-\underline{s}}$. Thus, the condition for $A^{*}>\underline{\underline{q}}\,\underline{\underline{A}}^{**}+\underline{q}\,\underline{A}^{**}+\bar{q}\,\bar{A}^{**}+\bar{\bar{q}}\,\bar{\bar{A}}^{**}$ is always fulfilled.

F. Proof of Proposition 5

We show in that – with geometric discounting – the right-hand side of condition (11) is smaller than the right-hand side of condition (8), while the left-hand sides are the same. Under geometric discounting, $\delta=\rho$, we have

$$\underline{\underline{q}}\left(\underbrace{\underline{\underline{A}}^{**}+\rho\,\underline{s}\,p(\underline{\underline{A}}^{**})\underline{d}}_{\le\underline{A}^{*}+\rho\,\underline{s}\,p(\underline{A}^{*})\underline{d}}\right)+\underline{q}\left(\underbrace{\underline{A}^{**}+\rho\,\bar{s}\,p(\underline{A}^{**})\underline{d}}_{\le\underline{A}^{*}+\rho\,\bar{s}\,p(\underline{A}^{*})\underline{d}}\right)+\bar{q}\left(\underbrace{\bar{A}^{**}+\rho\,\underline{s}\,p(\bar{A}^{**})d}_{\le\bar{A}^{*}+\rho\,\bar{s}\,p(\bar{A}^{*})d}\right)+\bar{\bar{q}}\left(\underbrace{\bar{\bar{A}}^{**}+\rho\,\bar{s}\,p(\bar{\bar{A}}^{**})d}_{\le\bar{A}^{*}+\rho\,\bar{s}\,p(\bar{A}^{*})d}\right)$$

$$\le\left(\underline{\underline{q}}+\underline{q}\right)\underline{A}^{*}+(\bar{q}+\bar{\bar{q}})\bar{A}^{*}+\delta\left(\left(\underline{\underline{q}}\,\underline{s}+\underline{q}\,\bar{s}\right)p(\underline{A}^{*})\underline{d}+\left(\bar{q}\,\underline{s}+\bar{\bar{q}}\,\bar{s}\right)p(\bar{A}^{*})d\right)$$

$$=q\,\underline{A}^{*}+(1-q)\bar{A}^{*}+\delta\left(q\,p(\underline{A}^{*})\underline{d}+(1-q)p(\bar{A}^{*})d\right)$$

Thus, the right-hand side of (11) is smaller than the right-hand side of (8).

(b) The proof is analogous to the proof of Proposition 3 in A.3.

References

Acemoglu, D., Aghion, P., Bursztyn, L., Hemous, D., 2012. The environment and directed technical change. Am. Econ. Rev. 102 (1), 131–166.
Allen, M.R., Ingram, W.J., 2002. Constraints on future changes in climate and the hydrologic cycle. Nature 419 (6903), 224–232.
Alterskjær, K., Kristjánsson, J.E., 2013. The sign of the radiative forcing from marine cloud brightening depends on both particle size and injection amount. Geophys. Res. Lett. 40 (1), 210–215.
Andreoni, J., Levinson, A., 2001. The simple analytics of the environmental kuznets curve. J. Public Econ. 80, 269–286.
Arrow, K.J., Cropper, M.L., Gollier, C., Groom, B., Heal, G.M., Newell, R.G., Nordhaus, W.D., Pindyck, R.S., Pizer, W.A., Portney, P.R., Sterner, T., Tol, R.S.J., Weitzman, M.L., 2014. Should governments use a declining discount rate in project analysis? Rev. Environ. Econ. Policy 8 (2), 145–163.
Aswathy, V.N., Boucher, O., Quaas, M., Niemeier, U., Muri, H., Mülmenstädt, J., Quaas, J., 2015. Climate extremes in multi-model simulations of stratospheric aerosol and marine cloud brightening climate engineering. Atmos. Chem. Phys. 15 (16), 9593–9610.
Barrett, S., 2014. Solar geoengineering's brave new world: thoughts on the governance of an unprecedented technology. Rev. Environ. Econ. Policy 8 (2), 249–269.
Brocas, I., Carrillo, J.D., 2000. The value of information when preferences are dynamically inconsistent. Eur. Econ. Rev. 44 (4–6), 1104–1115.
Carrillo, J.D., Mariotti, T., 2000. Strategic ignorance as a self-disciplining device. Rev. Econ. Stud. 67 (3), 529–544.
Cicerone, R., 2006. Geoengineering:encouraging research and overseeing implementation. Clim. Change 77, 221–226.
Convention on Biological Diversity (CBD), 2010. Decision Adopted by the Conference of the Parties to the Convention on Biological Diversity at its Tenth Meeting. UNEP/CBD/COP/DEC/X/33.
Corner, A., Pidgeon, N., 2010. Geoengineering the climate: the social and ethical implications. Environment 52 (1), 24–37.
Courbage, C., Rey, B., 2012. Optimal prevention and other risks in a two-period model. Math. Social. Sci. 63 (3), 213–217.
Cropper, M.L., Freeman, M.C., Groom, B., Pizer, W.A., 2014. Declining discount rates. Am. Econ. Rev. 104 (5), 538–543.
Crutzen, P.J., 2006. Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma? Clim. Change 77 (3), 211–220.
Davis, W., 2009. What does "Green" mean? Anthropogenic climate change, geoengineering, and international environmental law. Ga. Law Rev. 43 (3), 901–952.
Eeckhoudt, L., Huang, R.J., Tzeng, L.Y., 2012. Precautionary effort: a new look. J. Risk Insur. 79 (2), 585–590.
Ehrlich, J., Becker, G.S., 1972. Market insurance, self-insurance and self-protection. J. Polit. Econ. 80, 623–648.
Emmerling, J., Tavoni, M., 2017. Climate engineering and abatement: a 'flat' relationship under uncertainty. Environ. Resour. Econ., http://dx.doi.org/10.1007/s10640-016-0104-5.
Ferraro, A.J., Charlton-Perez, A.J., Highwood, E.J., 2014. A risk-based framework for assessing the effectiveness of stratospheric aerosol geoengineering. PLoS One 9 (2). (e88849 EP–).
Gardiner, S., 2010. Is arming the future with geoengineering really the lesser evil? Some doubts about the ethics of intentionally manipulating the climate

system. In: Gardiner, S., Jamieson, D., Caney, S. (Eds.), Climate Ethics: Essential Readings, Oxford University Press, Oxford.
Gerlagh, R., Liski, M., 2017. Consistent climate policies. J. Eur. Econ. Assoc.. (forthcoming).
Goeschl, T., Heyen, D., Moreno-Cruz, J., 2013. The intergenerational transfer of solar radiation management capabilities and atmospheric carbon stocks. Environ. Resour. Econ. 56 (1), 85–104.
Gollier, C., 2008. Discounting with fat-tailed economic growth. J. Risk Uncertain. 37 (2/3), 171–186.
Gollier, C., Treich, N., 2003. Decision-making under scientific uncertainty: the economics of the precautionary principle. J. Risk Uncertain. 27 (1), 77–103.
Goodell, J., 2010. How to Cool the Planet: Geoengineering and the Audacious Quest to Fix Earth's Climate. Houghton Mifflin, New York.
Groom, B., Hepburn, C., Koundouri, P., Pearce, D., 2005. Declining discount rates: the long and the short of it. Environ. Resour. Econ. 32 (4), 445–493.
Hale, B., Dilling, L., 2012. A charter for geoengineering (editorial). Nature 485, 415.
Heal, G., Millner, A., 2013. Discounting Under Disagreement. NBER Working Paper 18999.
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. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Jamieson, D., 1996. Ethics and intentional climate change. Clim. Change 33 (3), 323–336.
Jones, A., Haywood, J.M., Alterskjr, K., Boucher, O., Cole, J.N.S., Curry, C.L., Irvine, P.J., Ji, D., Kravitz, B., Egill?Kristjánsson, J., Moore, J.C., Niemeier, U., Robock, A., Schmidt, H., Singh, B., Tilmes, S., Watanabe, S., Yoon, J.-H., 2013. The impact of abrupt suspension of solar radiation management (termination effect) in experiment g2 of the geoengineering model intercomparison project (geomip). J. Geophys. Res.: Atmos. 118 (17), 9743–9752.
Keith, D.W., 2000. Geoengineering the climate: history and prospect. Annu. Rev. Energy Environ. 25 (1), 245–284.
Keith, D.W., Parson, E., Morgan, M.G., 2010. Research on global sun block needed now. Nature 463 (7280), 426–427.
Kimball, M., 1990. Precautionary saving in the small and in the large. Econometrica 58 (1), 53–73.
Klepper, G., Rickels, W., 2014. Climate engineering: economics prospects and considerations. Rev. Environ. Econ. Policy 8 (2), 270–289.
Klibanoff, P., Marinacci, M., Mukerji, S., 2005. A smooth model of decision making under ambiguity. Econometrica 73 (6), 1849–1892.
Koopmans, T.C., 1960. Stationary ordinal utility and impatience. Econometrica 28 (2), 287–309.
Korhonen, H., Carslaw, K.S., Romakkaniemi, S., 2010. Enhancement of marine cloud albedo via controlled sea spray injections: a global model study of the influence of emission rates, microphysics and transport. Atmos. Chem. Phys. 10 (9), 4133–4143.
Kravitz, B., Caldeira, K., Boucher, O., Robock, A., Rasch, P.J., Alterskjr, K., Karam, D.B., Cole, J.N.S., Curry, C.L., Haywood, J.M., Irvine, P.J., Ji, D., Jones, A., Kristjánsson, J.E., Lunt, D.J., Moore, J.C., Niemeier, U., Schmidt, H., Schulz, M., Singh, B., Tilmes, S., Watanabe, S., Yang, S., Yoon, J.-H., 2013. Climate model response from the geoengineering model intercomparison project (geomip). J. Geophys. Res.: Atmos. 118 (15), 8320–8332.
Kuebbeler, M., Lohmann, U., Feichter, J., 2012. Effects of stratospheric sulfate aerosol geo-engineering on cirrus clouds. Geophys. Res. Lett. 39 (23). (n/a).
Latham, J., 1990. Control of global warming? Nature 347, 339–340.
Long, J., Rademaker, S., Anderson, J., Benedick, R., Caldeira, K., Chaisson, J.e.a., 2011. Geoengineering: A National Strategic Plan for Research on the Potential Effectiveness, Feasibility, and Consequences of Climate Remediation Technologies. Task force on climate remediation Research, Bipartisan Policy Center.
Matthews, H.D., Caldeira, K., 2007. Transient climate-carbon simulations of planetary geoengineering. Proc. Natl. Acad. Sci. 104 (24), 9949–9954.
Moreno-Cruz, J.B., Keith, D.W., 2013. Climate policy under uncertainty: a case for solar geoengineering. Clim. Change 121 (3), 431–444.
Morrow, D., 2014. Ethical aspects of the mitigation obstruction argument against climate engineering research. Philos. Trans. R. Soc. A 372, 20140062.
Morrow, D., Kopp, R., Oppenheimer, M., 2009. Toward ethical norms and institutions for climate engineering research. Environ. Res. Lett. 4, 045106.
National Research Council, 2015. Climate Intervention: Reflecting Sunlight to Cool Earth. The National Academies Press, Washington, DC.
Nordhaus, W., 2011. The economics of tail events with an application to climate change. Rev. Environ. Econ. Policy 5 (2), 240–257.
Nordhaus, W.D., Sztorc, P., 2013. Dice 2013r: Introduction and User's Manual. URL ⟨http://www.econ.yale.edu/~nordhaus/homepage/documents/DICE_Manual_103113r2.pdf⟩.
Parson, E., Keith, D., 2013. End the deadlock on governance of geoengineering research. Science 339, 1278–1279.
Pindyck, R.S., 2013. Climate change policy: what do the models tell us? J. Econ. Lit. 51 (3), 860–872.
Quaas, J., Quaas, M.F., Rickels, W., Boucher, O., 2016. Regional climate engineering by radiation management: pCrerequisites and prospects. Earth's Future.
Rayner, S., Heyward, C., Kruger, T., Pidgeon, N., Redgwell, C., Savulescu, J., 2013. The oxford principles. Clim. Change 121, 499–512.
Ricke, K.L., Rowlands, D.J., Ingram, W.J., Keith, D.W., Granger Morgan, M., 2012. Effectiveness of stratospheric solar-radiation management as a function of climate sensitivity. Nat. Clim. Change 2 (2), 92–96.
Robock, A., 2008. 20 reasons why geoengineering may be a bad idea. Bull. At. Sci. 64 (2), 14–18.
Robock, A., 2012. Is geoengineering research ethical? Peace Secur. 30 (4), 226–229.
Robock, A., MacMartin, D.G., Duren, R., Christensen, M., 2013. Studying geoengineering with natural and anthropogenic analogs. Clim. Change 121 (3), 445–458.
Royal Society, 2009. Geoengineering the Climate – Science, Governance and Uncertainty. Vol. 10/09 of RS Policy document.
Schäfer, S., Irvine, P., Hubert, A., Reichwein, D., Low, S., Stelzer, H., Maas, A., Lawrence, M., 2013. Field tests of solar climate engineering. Nat. Clim. Change 3, 766.
Schneider, S.H., 1996. Geoengineering: could or should we do it? Clim. Change 33 (3), 291–302.
Schwartz, S.E., 2012. Determination of earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing. Surv. Geophys. 33, 745–777.
Stern, N., 2007. The Economics of Climate Change – The Stern Review. Cambridge University Press, Cambridge.
Storelvmo, T., Kristjansson, J.E., Muri, H., Pfeffer, M., Barahona, D., Nenes, A., 2013. Cirrus cloud seeding has potential to cool climate. Geophys. Res. Lett. 40, 178–182.
Stuart, G.S., Stevens, R.G., Partanen, A.-I., Jenkins, A.K.L., Korhonen, H., Forster, P.M., Spracklen, D.V., Pierce, J.R., 2013. Reduced efficacy of marine cloud brightening geoengineering due to in-plume aerosol coagulation: parameterization and global implications. Atmos. Chem. Phys. 13 (20), 10385–10396.
Traeger, C.P., 2009. Recent developments in the intertemporal modeling of uncertainty. Annu. Rev. Resour. Econ. 1, 261–285.
UNFCCC, 1992. United Nations Framework Convention on Climate Change. Available at ⟨http://unfccc.int/resource/docs/convkp/conveng.pdf⟩.
UNFCCC, 1997. Kyoto Protocol to the United Nations Framework Convention on Climate Change. Available at ⟨http://unfccc.int/resource/docs/convkp/kpeng.pdf⟩.
UNFCCC, 2015. Adoption of the Paris Agreement; Conference of the Parties twenty-first session Paris, 30 November–11 December 2015. Available at ⟨https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf⟩.
Victor, D.G., 2008. On the regulation of geoengineering. Oxf. Rev. Econ. Policy 24 (2), 322–336.
Wang, J., Li, J., 2015. Precautionary effort: another trait for prudence. J. Risk Insur. 82 (4), 977–983.
Weitzman, M., 2009. On modeling and interpreting the economics of catastrophic climate change. Rev. Econ. Stat. 91 (1), 1–19.
Weitzman, M., 2011. Fat-tailed uncertainty in the economics of catastrophic climate change. Rev. Environ. Econ. Policy 5 (2), 275–292.
Williamson, P., Turley, C., 2012. Ocean acidification in a geoengineering context. Philos. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci. 370 (1974), 4317–4342.
Züern, M., Schäfer, S., 2013. The paradox of climate engineering. Glob. Policy 4 (3), 266–277.