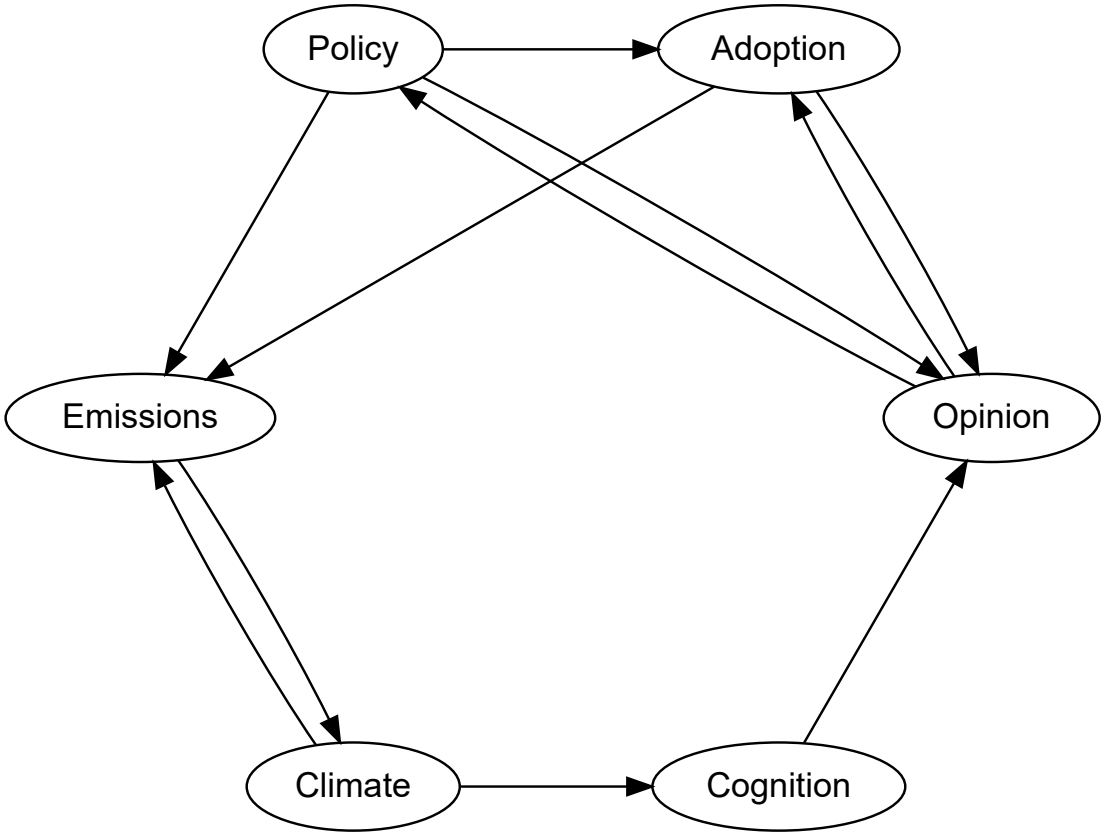
Climate-Social System Model Documentation

Model Overview

There are six primary components in the model, with linkages between components shown below. Opinions about climate policy (*Opinion Component*) determine both individual adoption of emissions-reducing behavior (*Adoption Component*) and collective action (*Policy Component*). Adoption and policy togther determine the level of greenhouse gas emissions (*Emissions Component*), which in turn affect the climate system (*Climate Component*). Observation of climate change can feedback to affect opinions on climate policy (*Cognition Component*).



Opinion Component

Population is divided into three types based on opinion (Op) regarding climate policy - Supporting, Neutral, and Opposed (i.e. $\mathbf{Op} = [S, N, O]$). Movement between adjacent types is determined by up to three things: a social pressure or persuasive force based on interactions between groups given the group network structure ($Force$), evidence derived from one's own experience of weather (Ev), and a feedback from policy to opinion driven by individuals updating their information about the social norm based on institutional (i.e. policy) change ($PolOp$).

Persuasive forces are governed by two parameters, the persuasive force of neutral on opinionated (the "weak persuasive force", F_N), and the persuasive force of opinionated on neutral (the "strong persuasive force", $F_{S,O}$). There is also a "credibility-enhancing display" feedback allowed from the Adoption Component. This allows the persuasive effect of climate policy supporters to be larger depending on the level of adoption of individual sustainable behaviors among supporters ($A_{t-1,S}$), compared to other opinion groups. The size of the effect is controlled by the CED parameter (ξ). Therefore, the force matrix describing the persuasive effect felt by each group (rows) from each other group (colums), conditional on encounter is given by:

$$\mathbf{Force_t} = \begin{bmatrix} 0 & F_N & F_{S,O} \\ F_{S,O} + \xi(A_{t-1,S} - A_{t-1,N}) & 0 & F_{S,O} \\ F_{S,O} + \xi(A_{t-1,S} - A_{t-1,O}) & F_N & 0 \end{bmatrix}$$

The force felt by each group depends on the interaction of the force matrix (probability of conversion, conditional on contact) with the social network (probability of contact). The social network is related to the distribution of opinions in the population, but is not entirely determined by it. The network homophily parameter (θ) allows for preferential interaction between opinions of the same type, reflecting that there may be social factors that cause contact or information sharing to disproportionately happen between rather than within groups. In the interests of parsimony, θ gives the differential probability of encounter with ones own group and it is assumed that other encounters are split equally with the other two groups. Therefore $\theta = \frac{1}{3}$ is a fully mixed population and $\theta = 1$ is a fully separated population.

The social network defining probability of intereaction for each group (rows) with each other group (columns) is given by:

$$\mathbf{Network_t} = \begin{bmatrix} \theta S_t & \frac{1-\theta}{2} N_t & \frac{1-\theta}{2} O_t \\ \frac{1-\theta}{2} S_t & \theta N_t & \frac{1-\theta}{2} O_t \\ \frac{1-\theta}{2} S_t & \frac{1-\theta}{2} N_t & \theta O_t \end{bmatrix}$$

The social force is given by element-wise multiplication of the network and force matrices
$\mathbf{SocialForce_t} = \mathbf{Network_t} * \mathbf{Force_t}$

In addition to social persuasion, transition between opinion groups is also determined directly by the experience of climate chance through the perception of weather anomalies. This evidence effect (Ev) is allowed to differ by opinion group (reflecting biased assimilation or other forms of motivated reasoning), can take positive or negative values (depending on whether the perceived weather anomaly is warm or cold), and is scaled by the "evidence effectiveness" parameter, (η) which determines how strongly perceived weather anomalies affect opinion. More details on how the evidence based on climate change is calculated is in the Cognition Component.

Finally, we also allow for a feedbacks allowing changes in policy to influence opinion via the "Expressive Force of Law" feedback documented in the legal literature.

The feedback from policy to opinion operates by providing evidence on the population-wide norm (as oppposed to the norm experienced by individuals which depends on the social network). The strenght of this effect depends on the change in policy last period ($\Delta Policy_{t-1}$) and a scaling parameter κ:

$$PolOp_t = \kappa \Delta Policy_{t-1}$$

Note that this effect is the same for all opinion groups and that the sign can be either positive or negative, depending on whether the policy change last period increased (positive) or decreased (negative) pro-climate policies.

The transition probability matrix, restricting transitions to neighboring opinion groups, is given by:

$$\mathbf{P_t} = \begin{bmatrix} 1 - P_{1,2} & \sum_{j=2,3} SocialForce_{1,j} - \eta Ev_{1,t-1} - PolOp_t & 0 \\ SocialForce_{2,1} + \eta Ev_{2,t-1} + PolOp_t & 1 - (\sum_{j=1,3} SocialForce_{2,j}) & SocialForce_{2,1} - \eta Ev_{2,t-1} - PolOp \\ 0 & \sum_{j=1,2} SocialForce_{3,j} + \eta Ev_{3,t-1} + PolOp_t & 1 - P_{3,2} \end{bmatrix}$$

The new distribution of opinion in time t depends on the the previous distribution of opinion and transition probabilities:

$$\mathbf{Op_t} = \mathbf{Op_{t-1}} * \mathbf{P_t}$$

Policy Component

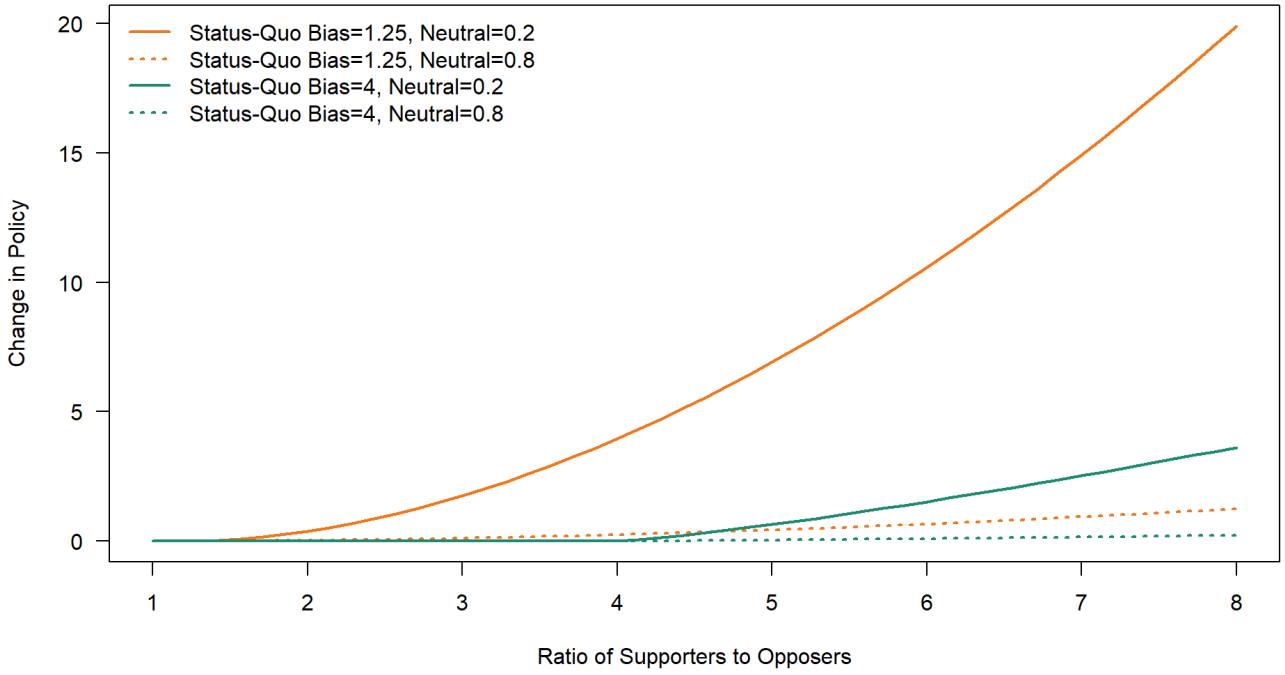
Collective policy is summarized by a single value, which can be positive or negative, and can be thought of as the magnitude of the tax on carbon emissions (or subsidy of fossil fuels if negative). Policy is cumulative, with the change in policy each period determined by the distribution of opinion and a single parameter, ν, which captures the magnitude of institutional bias towards the status quo.

If the number of supporters is greater than the number of opposers (i.e. $S_t > O_t$) the the change in policy in period t is given by:

$$\Delta Policy_t = \begin{cases} 0 & \text{if } \frac{S_t}{O_t} < \nu \\ (1 - N_t)^2 * (\frac{S_t}{O_t} - \nu)^{(1+\frac{1}{\nu})} & \text{otherwise} \end{cases}$$

The status quo bias parameter does two things. Firstly it creates a threshold that majorities of opinion must overcome, before they are able to influence policy. Secondly it affects the rate at which policy responds to larger opinion majorities beyond the threshold. At the lower bound of $\nu = 1$, simple majorities are enough to effect change and the rate of change is a quadratic in the size of the majoritiy. As ν increases, the opinion threshold becomes larger and the rate of increase tends towards linear (i.e. $(1 + \frac{1}{\nu}) \underset{\nu \to \infty}{\longrightarrow} 1$). The population of people without strong opinions on climate policy (N_t) creates inertia in the policy response, so that the policy response is smaller if a large fraction of people do not have strong opinions about policy.

The functional form with two values of ν and two values of N_t are given below:



The formulation is symmetric in the opposite case that $O_t > S_t$, with the change in policy then being negative rather than positive. Policy is limited by an arbitrary maximimum value of 300.

An interest group feedback effect allows the the status quo bias to change as function of previous policy. It can be either a postive or a negative feedback. A positive feedback captures the ways in which policy change can establish new business or public interests that can be mobilized to advocate for more poilicy. A negative feedback reflects the possibility that small policy change can mobilize existing powerful interests opposed to

policy change to lobby against further change. The feedback therefore has asymmetrical or directional effect on the status-quo bias - increasing it to resist chnage in one direction while decreasing it in the other direction.

The size of the interest group feedback effect (Ig) depends on the feedback parameter ζ and on the average value of policy over some previous time window (w), set by default to 10 years:

$$Ig = \zeta * \sqrt{\frac{|\overline{Policy_w}|}{300}}$$

Where $|\overline{Policy_w}|$ is the mean absolute value of policy over the previous w year window. The square-root functional form allows for largest changes at small policy magnitudes (for instance by the creation of new industries) that saturates at larger values. Division by the maximum policy value (300) limits the maximum feedback effect to ζ. The sign of Ig is either positive or negative depending on the sign of ζ.

The feedback effect is applied to the status-quo bias parameter differently depending on the sign of policy over the window w. If past policy was pro-climate (i.e. positive sign), the status quo bias against more pro-climate changes (ν_{pro}) becomes $\nu_{pro} = \nu - Ig$ and the status quo bias against reversing that previous policy becomes ($\nu_{opp} = \nu + Ig$). Note that if ζ and therefore Ig is positve, this decreases the status-quo bias against more climate policy, whereas if ζ is negative this increases the status-quo bias. The effect is symmetric if mean policy over w was negative instead of positive.

Adoption Component

Individual behavior to reduce emissions is represented by a choice to adopt or not adopt a single sustainable behavior that reduces individual emissions. It can be thought of as a representitive composite of all the actions individuals can take to reduce emissions. The adoption decision is determined by two things: 1) the social norm around adoption / non-adoption determined by the social network for each opinion group and 2) perceived behavioral control, the opposite of adoption costs, which are affected by the total number of adopters, policy, and opinion.

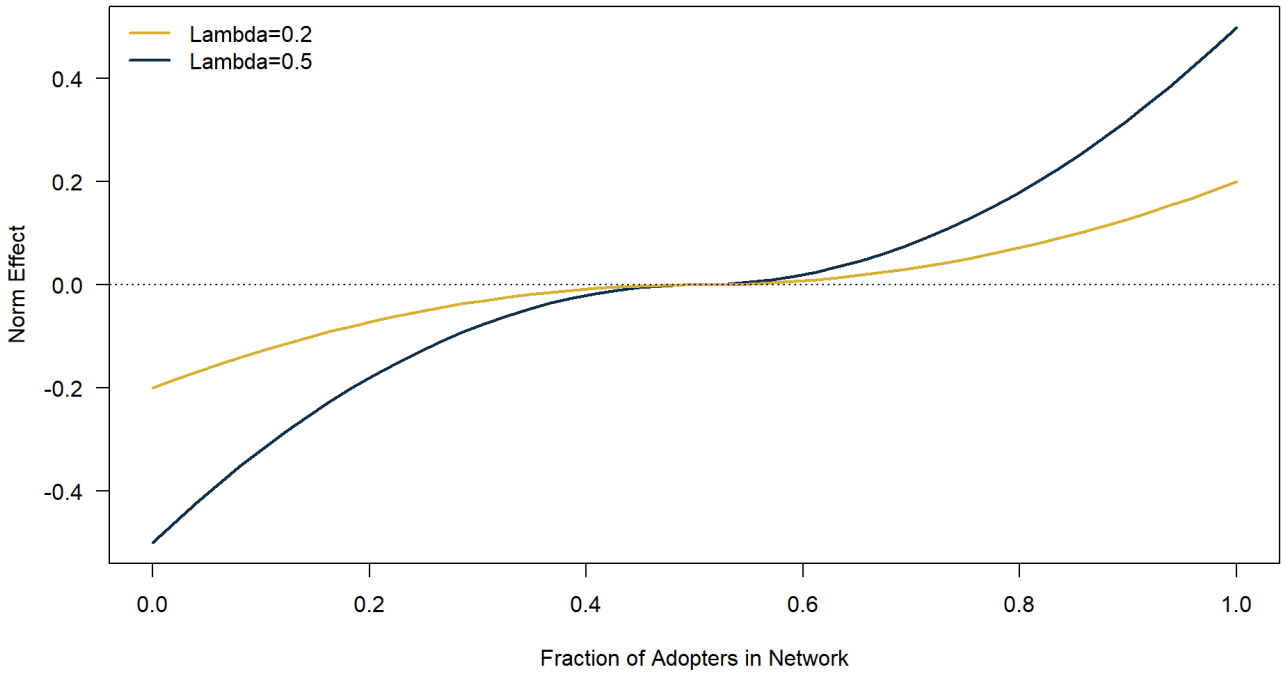
The fraction of adopters within each opinion group is given by the vector $\mathbf{A}$. The fraction of adopters within the social network of opinion group i is given by:

$$Frac_i = \sum_{j=1,2,3} Network_{i,j} * A_j$$

The social norm around adoption felt by members of each opinion group depends on the fraction of network adopting and a "norm effectiveness" scaling parameter (λ) which determines the strength of the social norm effect:

$$Norm_i = \begin{cases} \lambda * (1 - 4Frac_i + 4Frac_i^2) & \text{if } Frac_i > 0.5 \\ -\lambda * (1 - 4Frac_i + 4Frac_i^2) & \text{if } Frac_i <= 0.5 \end{cases}$$

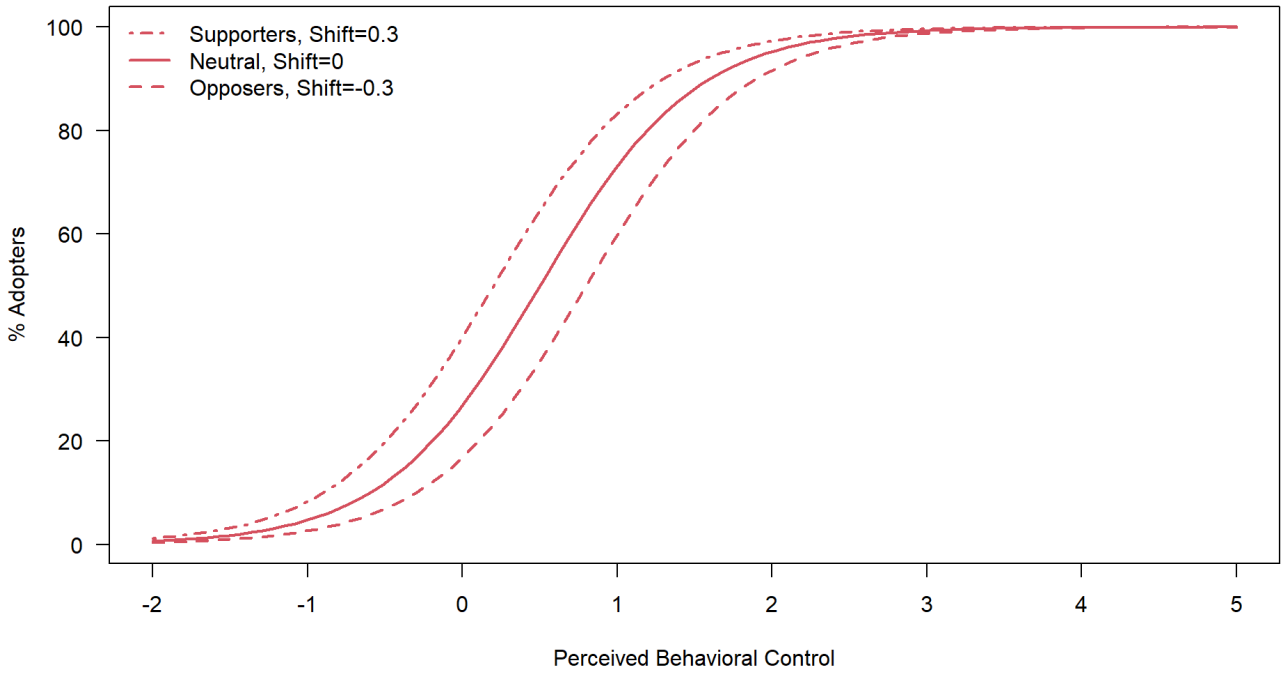
This functional form has the desireable properties that $Norm_i = 0$ for $Frac_i = 0.5$ and that increases in $Frac_i$ have a larger effect on the norm towards the extreme values of 0 and 1. The functional form is shown below for two values of λ:



Adoption also depends logistically on perceived behavioural control (pbc), conceptually similar to the inverse of adoption cost. The logistic function giving the fraction of adopters as a function of pbc is determined by 2 parameters, pbc_{mid} which gives the value of pbc with 50% adoption, and pbc_{steep}, which controls how steeply adoption responds to changes in pbc away from the midpoint. Opinion about climate change can also affect the adoption curve, by shifting the midpoint of the adoption curve, meaning that the same value of pbc results in different adoption rates in the three different opinion groups. This parameter, $pbc_{shift,i}$ is specific to opinion group. The fraction of adopters in opinion group i in time t is given by:

$$A_{it} = Norm_{it} + \frac{1}{(1 + e^{-pbc_{steep} * (pbc_t - (pbc_{mid} - pbc_{shift,i}))})}$$

Example logistic curves for three opinion groups, assuming $Norm_{it} = 0$, are shown below:



PBC each time period depends on two things: the total number of adopters in the previous time period ($N_{t-1} = \sum_i A_{i,t-1} * Op_{i,t-1}$) and policy. The effect of number of adopters captures a generic endogenous technical change (etc) effect, where more people using a technology lowers the cost for others, through mechanisms such as economies of scale, network effects, and technical or social learning. This effect is parameterized as a logistic, with three parameters: etc_{total} gives the maximum effect of etc, etc_{mid} gives the number of adopters required for 50% realization of the total etc effect, and etc_{steep} gives the steepness of the response.

PBC at time t is therefore given by:

$$pbc_t = pbc_0 + \frac{etc_{total}}{(1 + e^{-etc_{steep} * (N_{t-1} - etc_{mid})})} + pbc_{pol,t}$$

The term $pbc_{pol,t}$ allows policy to affect adoption costs (pbc), for example by making sustainable options more or less expensive relative to an alternative. This is directional (i.e. positive, emissions-reducing policy raises pbc, increasing adoption while negative policy reduces pbc). This effect is assumed to be linear in policy, up to a maximum policy effect, ($maxpbc_{pol}$):

$$pbc_{pol,t} = \begin{cases} Policy_t * 0.1 & \text{if } |Policy_t * 0.1| < maxpbc_{pol} \\ maxpbc_{pol} & \text{if } Policy_t * 0.1 > maxpbc_{pol} \\ -maxpbc_{pol} & \text{if } Policy_t * 0.1 < -maxpbc_{pol} \end{cases}$$

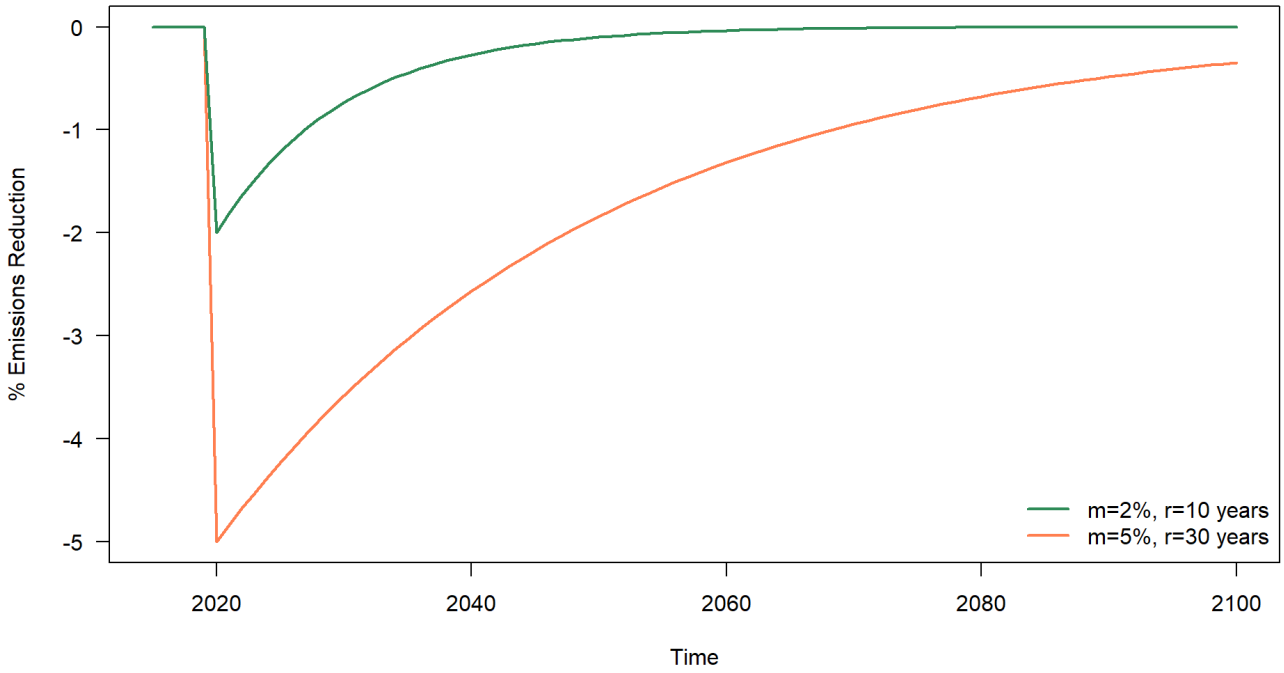
Emissions Component

Global emissions in the absence of climate policy (i.e. positive values of $Policy$) are given by RCP 7.0 and are denoted BAU. This is a relatively high emissions scenario in which global emissions double by 2100 and warming approaches 4° above pre-industrial temperatures by 2100.

Reduction in emissions comes from two sources, adoption of pro-climate behaviors by individuals and collective policy action. Individual action is assumed to produce instantaneous emissions reduction, but not to

have any persistent (i.e. cumulative) effect. Therefore mitigation resulting from individual adoption of sustainable behaviors in year t is given by $M_{ind,t} = \pi N_t$ where π is a scaling factor describing the mitigation effectiveness of individual adoption.

Policy also affects emissions directly. These emissions reductions are modeled as persistent but not permanent. The effect of policy today on emissions in the future is modeled using an exponential decay function, parameterized using two variables: the contemporaneous effect of policy today on emissions, m; and the lifetime of contemporaneous investments in mitigation, r. Both m and r are allowed to increase with increasing policy, reflecting the fact that more stringent climate policy might produce both larger and longer-lasting emissions reductions, in the form of investments on longer-lived infrastructure. The time-path of emissions reductions for two combinations of m and r are given below:



Both m and r increase with increasing policy. The change in m with policy is given by

$$m_t = \begin{cases} \gamma_t * \frac{log(Policy_t)}{log(300)} & \text{if } Policy_t < 300 \\ \gamma_t & \text{otherwise} \end{cases}$$

Where γ is a parameter describing the maximum contemporaneous reduction in emissions and 300 comes from the arbitrary maximum value of $Policy$ (Policy Component).

A learning-by-doing effect in which mitigation technology costs change as a function of installed capacity is modeled by allowing the γ_t parameter to change as a function of cumulative, policy-induced mitigation in the previous time period (i.e. $M_{pol,t-1}$, see below for M_{pol} definition), and a learning-by-doing parameter (lbd) that gives the fraction reduction in costs for each doubling (here definied relative to an initial γ_0 value):

$$\gamma_t = \gamma_0 * (1 + lbd)^{log_2(M_{pol,t-1}/\gamma_0)}$$

Note that the exponent gives the number of doublings relative to the initial γ value. This is a representation of a "single-factor learning curve" described by Rubin et al. (2015). Their review gives ranges for the lbd parameter between 0 and 30%. γ_0 is the "Max Mitigation Rate" parameter and is calibrated based on data

from Andersson (2019) on the effectiveness of the Swedish carbon tax.

The half-life of mitigation investment, r, is assumed for simplicity to be linear in $Policy$ from an initial value, r_0 to a maximum value, δ:

$$r_t = min(r_0 * (1 + \frac{Policy_t}{10}), \delta)$$

The policy-induced mitigation in time t is given by the sum of the current effects of mitigation in all previous time periods:

$$M_{pol,t} = \sum_{i=1}^{t} m_i e^{-(t-i)*r_i}$$

Therefore, emissions in period t are given by:

$$E_t = BAU_t(1 - M_{pol,t})(1 - M_{ind,t}) = BAU_t(1 - M_{pol,t})(1 - \pi N_t)$$

By default, the dynamics represented in the Opinion, Adoption, Policy, and Emissions components are based on studies from relatively wealthy countries. Accordingly, mitigation rates are applied to the subset of emissions from OECD countries. To simulate global temperatures and therefore the feedback from the climate system to opinion (Congition Component), mitigation rates in the rest of the world also need to be modeled. We model this simply as the lagged mitigation rates in the OCED, parameterized with the lag parameter, L, giving the number of years lag:

$$E_{Tot,t} = E_t + BAU_{O,t} * (1 - \frac{(BAU_{t-L} - E_{t-L})}{BAU_{t-L}})$$

Where $BAU_{O,t}$ is business as usual emissions in the outside region. The model can be collapsed to a single-region model by setting $L = 0$

Finally, there is a feedback allowed from global temperature to emissions. This reflects the fact that climate change itself may affect the expected level of emissions through impacts on economic productivity, the energy intensity of production, or the carbon intensity of energy. Following Woodard et al. (2019) this is parameterized as a % reduction in emissions that increases linearly in temperature:

$$E_{Tot,t} = (1 + (tef * T_{t-1})E_{Tot,t}$$

Where tef is the temperature emissions feedback that has a central value of -0.031 based on Woodard et al. (2019).

Climate Component

The climate model is based on the DICE model, annualized using formulas from Cai and Lontzek (2019). This is a three box carbon-cycle model (atmopsphere, upper ocean, and lower ocean) that tracks both atmopspheric and ocean temperature. Details are available in: Cai, Y. and Lontzek, T. S., 2019. "The Social Cost of Carbon with Economic and Climate Risks", Journal of Political Economy, 127 (6), pp.2684-2736 (https://www.journals.uchicago.edu/doi/full/10.1086/701890?mobileUi=0#)

The Cai and Lontzek model is initialized in 2005. For the default model run, we re-initialize using carbon mass and radiative balances in 2020 by running the model using global emissions observed 2005-2020. Following the DICE 2016 model,forcing from non-CO_2 greenhouse gases is parameterized as a single "exogenous forcing" term that is added to forcing from CO_2. This forcing is scaled using the relationship:

$$ExF_t = ExF_{BAU,t} * (1 - 0.49 * f_{CO2,t})$$

where ExF_t is the exogenous forcing term in time t, $ExF_{BAU,t}$ is the business as usual exogenous forcing term, and $f_{CO2,t}$ is the fractional reduction in CO_2 emissions below the RCP7 baseline in year t defined as $f_{CO2} = \frac{(BAU_t - E_{Tot,t})}{BAU_t}$. The 0.49 scaling term comes from the relationship between reductions in CO_2 emissions and reductions in CH_4 and N_2O emissions below RCP7 values observed across other RCP-SSP scenarios.

Cognition Component

If $\eta > 0$ then experience of climate change is allowed to affect opinion about climate policy. In particular, the climate system provides evidence (Ev) in the form of perceived temperature anomalies. Individuals experience temperature anomalies based both on the mean climate state (ΔT_t, caluclated in the Climate Component) and a random draw of weather, w_t. Weather anomlies are centered on zero and are a random realization of a spectral decomposition of a 500-year pre-industrial climate model run. This preserves the temporal dependnce of weather anomalies, capturing natural cycles such as El Nino, which could affect the perception of climate change over short timescales.

Two forms of imperfect cognition are represented: shifting baselines and biased assimilation. Shifting baselines reflect the fact that the evalutation of temperatures may change over time as older conditions are forgotten. The perceived anomaly is therefore given by:

$$Anomaly_t = \begin{cases} \Delta T_t + w_t & \text{if } Base = Fixed \\ \Delta T_t + w_t - \sum_{i=8}^{2} \beta_i (\Delta T_{t-i} + w_{t-i}) & \text{if } Base = Shifting \end{cases}$$

In other words, in the absence of shifting baselines, the perceived anomaly is simply experienced change in weather since pre-industrial. If baselines shift, however, the anomlay is perceived only relative to a baseline based on a weighted average of the last 2-8 years, Parameterization of the shifiting baseline, including the window and the weighting of experienced temperatures within that window, comes from Moore et al. (2019).

A biased assimilation effect is a form of motivated reasoning that allows selective incorporation of evidence into one's beliefs. This is modeled by allowing temperature anomalies to constitute different levels of evidence (Ev) for different opinion groups. The strength of this effect is given by the parameter μ which up- or down-weights the temperature anomaly, depending on the sign of the anomaly and the opinion group. The effect is fixed at zero for the neutral opinion group.

Therefore, experiential evidence for climate change for those that already support climate change policy is given by:

$$Ev_{S,t} = \begin{cases} (1 + \mu) Anomaly_t & \text{if } Anomaly_t > 0 \\ (1 - \mu) Anomaly_t & \text{if } Anomaly_t < 0 \end{cases}$$

For those neutral on climate policy $Ev_{N,t} = Anomaly_t$, and for those opposing climate policy:

$$Ev_{O,t} = \begin{cases} (1 - \mu) Anomaly_t & \text{if } Anomaly_t > 0 \\ (1 + \mu) Anomaly_t & \text{if } Anomaly_t < 0 \end{cases}$$

If $\eta > 0$ then this evidence affects opinions about climate policy, as part of the Opinions Component.