Ambiguous Tipping Points

Derek Lemoine† & Christian Traeger§

† Department of Economics, University of Arizona
McClelland Hall 401, 1130 E Helen St, Tucson, AZ, 85721-0108, USA
dlemoine@email.arizona.edu
§ Department of Agricultural & Resource Economics, University of California, Berkeley
207 Giannini Hall #3310, Berkeley, CA 94720-3310, USA
traeger@berkeley.edu

August 2015

Abstract: We analyze the policy implications of aversion to Knightian uncertainty (or ambiguity) about the possibility of tipping points. We demonstrate two channels through which uncertainty aversion affects optimal policy in the general setting. The first channel relates to the policy’s effect on the probability of tipping, and the second channel to its differential impact in the pre- and post-tipping regimes. We then extend a recursive dynamic model of climate policy and tipping points to include uncertainty aversion. Numerically, aversion to Knightian uncertainty in the face of an ambiguous tipping point increases the optimal tax on carbon dioxide emissions, but only by a small amount.

JEL Codes: Q54, D90, D81, C61

Keywords: tipping point, ambiguity, Knightian uncertainty, threshold, regime shift, climate, hazard, integrated assessment, dynamic programming, social cost of carbon, carbon tax

*We are grateful for comments from Larry Karp, Ujjayant Chakravorty, and participants at a number of seminars, including the 2011 NBER Environmental and Energy Economics Summer Institute and the 2015 workshop on “Thresholds, Tipping Points and Random Events in Dynamic Economic Systems” at the University of Tennessee.
1 Introduction

Tipping points confront policy makers with the possibility of regime shifts. Policymakers usually have a thin or even non-existent record of past regime shifts from which to gauge the possibility of triggering such a shift in the future. For instance, European countries deciding whether to tolerate the bankruptcy of a member state must consider the unfamiliar chance of tipping into a new regime of high bond yields and further crises. Political elites deciding whether to expropriate resources must consider the chance of triggering a rare mass uprising. And, in our example of primary interest, policymakers deciding how to regulate greenhouse gas emissions must consider the unexperienced chance of irreversibly tipping the planet into a less favorable climate system.

Many agents appear to dislike Knightian uncertainty (Camerer and Weber 1992), and normative models of decision-making under uncertainty allow for such aversion (Traeger 2010, Cerreia-Vioglio et al. 2011, Etner et al. 2012, Gilboa and Marinacci 2013). Our policymaker demonstrates an aversion to low-confidence Bayesian priors in the framework of Traeger (2010), which relates closely to the “smooth” ambiguity model of Klibanoff et al. (2009).

We analyze how aversion to Knightian uncertainty alters optimal policy in the presence of tipping points with unknown triggers. We numerically solve for the optimal tax on carbon dioxide emissions in the face of deeply uncertain climate tipping points.

Lemoine and Traeger (2014) show that potential tipping points affect the optimal level of a policy control through two channels. We show how aversion to (Knightian) uncertainty about the threshold’s location changes each of these channels. First, the endogenous probability of tipping implies that the present policy influences the chance of

---

1 We use the terms “Knightian uncertainty”, “deep uncertainty”, “uncertainty”, and “ambiguity” interchangeably to describe a situation where the underlying probabilities are not known. We allow our decision-maker to exhibit higher aversion to situations with unknown probabilities. The smooth ambiguity model of Klibanoff et al. (2005, 2009) relies on subjective ambiguous priors over objective lotteries. By detaching ambiguity aversion from the particular hierarchical order of priors in the smooth ambiguity model, the framework of Traeger (2010) allows us to separate the ambiguous probability of triggering a tipping point from stochastic temperature shocks whose distribution is objectively known.

2 Gjerde et al. (1999), Keller et al. (2004), and Lontzek et al. (2015) also numerically analyze the implications of endogenous climate tipping points. Whereas they model tipping points as directly reducing utility or output, we follow Lemoine and Traeger (2014) in modeling tipping points as directly shifting the underlying dynamics of the climate system. The implications for utility and output then depend on how the policymaker responds to the altered climate dynamics. van der Ploeg (2014) also follows Lemoine and Traeger (2014) in modeling a tipping point as changing the dynamics of the carbon cycle.

3 Our analysis of uncertainty aversion corresponds to an analysis of risk aversion in a setting in which tipping points are the only stochastic element and Epstein-Zin preferences disentangle risk aversion from intertemporal consumption smoothing motives.
tipping. Lemoine and Traeger (2014) call this channel the marginal hazard effect (MHE) and show that it leads the optimal policy to reduce the likelihood of a harmful tipping point. We show that uncertainty aversion further increases the marginal hazard effect if the tipping hazard is small (and decreases the MHE if the probability of tipping is high). A rule of thumb is that uncertainty aversion changes the MHE in a direction that reduces Knightian uncertainty. In our climate change example, the annual probability of a tipping point is small and reducing emissions makes triggering a tipping point less likely. Uncertainty aversion increases the contribution of the MHE to the optimal tax on carbon dioxide emissions.

Second, the marginal effect of today’s policy on future welfare depends on whether a tipping point happens to occur. Lemoine and Traeger (2014) call this channel the differential welfare impact (DWI) because it is proportional to the difference in the welfare impact of the control in the pre- and post-tipping regimes. If the marginal welfare benefit of stronger policy is greatest in the post-tipping regime, then uncertainty aversion tends to strengthen policy through this second channel. In our climate example, tipping into a runaway climate may imply that strengthening policy generates higher payoffs in the post-tipping world. However, the endogeneity of the tipping hazard increases the value from having strengthened the policy if the world remains in the pre-threshold regime, which can outweigh the increased harm of carbon dioxide emissions in the post-threshold world (with a runaway climate). Our numeric application finds that uncertainty aversion makes the DWI reduce the optimal tax on carbon dioxide emissions. Uncertainty aversion increases the welfare cost of potential future tipping points (through the MHE) and so particularly increases the marginal welfare impact of the policy in the pre-threshold regime. However, the reduction in the DWI is small in comparison to the effect of uncertainty aversion on the marginal hazard effect. Overall, uncertainty aversion increases the optimal tax on carbon dioxide emissions.

Our work sits at the intersection of four literatures. First, a primarily theoretical literature investigates the implications of uncertainty aversion for optimal savings and portfolio allocations. Our DWI contribution relates closely to a self-insurance motive identified in these settings (Gollier 2011, Alary et al. 2013). Our MHE contribution relates closely to a self-protection motive (Snow 2011, Alary et al. 2013, Maccheroni et al. 2013). Traeger (2011) relates these uncertainty aversion effects to the precautionary savings arguments under risk. Our framework differs in two ways from this literature. First, it is more complex because self-insurance and self-protection interact (also over time), and because our discrete regime shift complicates the trade-off for self-insurance.
Second, our setting is simpler in that mitigation effort only affects the uncertain tipping hazard, whereas much of this literature has intertwined self-protection against uncertainty and standard risk. For the MHE effect, our findings are closest to Snow (2011), who finds that uncertainty aversion always increases self-protection, which holds in our setting under an approximation for small to moderately large hazard rates. As Alary et al. (2013) explain, this result arises in our setting because mitigation effort reduces uncertainty. The self-insurance motive in our setting, captured by the DWI, differs more significantly from the analysis in this literature because of our discrete regime shift.

Second, the literature has considered the implications of Knightian uncertainty in the context of climate change. In a stylized two-period example, Lange and Treich (2008) show that aversion to Knightian uncertainty about damages from climate change reduces optimal emissions. Millner et al. (2013) analyze the implications of scientific ambiguity about the effect of greenhouse gas emissions on global temperature. They consider how uncertainty aversion affects the value of exogenously defined emission policies. They find that the effect is analytically ambiguous but numerically works to increase the value of emission policies that limit warming. Other papers use Knightian uncertainty to motivate robust control approaches to emission policy (Hennlock 2009, Anderson et al. 2014, Rudik 2015). Our work relates most closely to Jensen and Traeger (2013, 2014), who analytically and numerically analyze the effects of Epstein-Zin preferences and smooth ambiguity aversion on optimal emission policy in a setting without tipping points.

Third, our numerical exploration of uncertainty aversion in the context of tipping points extends a recent literature that uses recursive dynamic models to explore the implications of uncertainty and learning for the optimal tax on carbon dioxide emissions. This literature has analyzed the implications of uncertainty about warming (Kelly and Kolstad 1999, Leach 2007, Jensen and Traeger 2013, Hwang et al. 2014, Kelly and Tan 2014), about economic growth (Jensen and Traeger 2014), about damages from climate change (Crost and Traeger 2013, 2014, Cai et al. 2013, Rudik 2015), and about tipping points under standard risk (Lemoine and Traeger 2014).

Finally, the environmental economics literature has a long history of discussing theoretical aspects of pollution-triggered regimes shifts. Clarke and Reed (1994) discuss a regime shift triggered by a stochastic process. Heal (1984) and Tsur and Zemel (1996) introduce epistemological uncertainty where the policymaker does not know the underlying threshold, as in our setting. Tsur and Zemel (2009) and de Zeeuw and Zemel (2012) emphasize that endogenous tipping risk can make the policymaker more cautious, and Nævdal (2006) analyzes the case where the tipping point’s trigger is a combination of dif-

The next section describes the theoretical setting and the determinants of optimal policy. Section 3 analyzes how uncertainty aversion affects the value of increasing a control. Section 4 describes our recursive dynamic model of climate policy, ultimately based on the benchmark DICE-2007 integrated assessment model of Nordhaus (2008). Section 5 reports our numerical results. Section 6 concludes.

2 Setting

This section explains our representation of tipping points as a regime shift triggered with an endogenous probability. Our policymaker faces an intertemporal trade-off between present benefits and future damages. We show how the presence of regime shifts changes this trade-off.

2.1 The tipping point

We analyze optimal policy in the face of irreversible shifts in system dynamics. These regime shifts are triggered upon crossing a threshold in the state space. The policymaker does not know the precise location of the threshold, has a subjective prior over the possible threshold locations, and learns about its location by observing whether tipping occurs. Our setting embeds the tipping points of Lemoine and Traeger (2014) into a setting with Knightian uncertainty.

The hazard rate is the probability of tipping into a new regime within a given period. This hazard rate is a function of prior beliefs about possible threshold locations and the set of possible threshold locations that the system passes through in a given period. The policymaker controls the tipping probability by controlling movement through the state space. This control is imperfect because, first, inertia in the system prevents full and immediate control of the system states and, second, because the system is subject to stochastic shocks. Our policymaker learns that thresholds that have not triggered a tipping point in the past will be safe in the future, and he updates his Bayesian prior governing threshold locations accordingly.

The policymaker is fully rational and forward-looking, and he incorporates the value of learning into his decisions. He solves an infinite-horizon dynamic optimization problem. We denote the vector of system states by $S_t$. In our climate change example, this vector
will include capital, temperature, carbon dioxide, and time (to account for exogenous nonstationary processes). The value function $V_\psi(S_t)$ characterizes the maximal expected welfare that the policymaker can derive from the optimally controlled system. The parameter $\psi$ indicates the regime: $\psi = 0$ indicates the pre-tipping world, whereas $\psi = 1$ indicates the post-tipping world. In our climate change example, the carbon dioxide and temperature dynamics will differ between the pre- and post-tipping regimes.

The policymaker’s utility $u(x_t, S_t)$ in a period depends on the system’s state and on the control vector $x_t$. The optimal post-tipping policy and value function solve the standard dynamic programming problem:

$$
V_1(S_t) = \max_{x_t} \left\{ u(x_t, S_t) + \beta_t \int V_1(S_{t+1}) d\mathbb{P} \right\} \\
\text{s.t. } S_{t+1} = g_1(x_t, \epsilon_t, S_t) \\
x_t \in \Gamma(S_t).
$$

The function $g_1(x_t, \epsilon_t, S_t)$ characterizes the next period’s state as a function of this period’s control vector $x_t$, the stochastic shock vector $\epsilon_t$, and this period’s state vector $S_t$. The set $\Gamma(S_t)$ constrains the control vector $x_t$. The probability measure $\mathbb{P}$ characterizes the distribution of the independently and identically distributed stochastic shocks $\epsilon_t$. The Bellman equation (1) is a functional equation for the value function $V_1$. Its solution determines the maximal value that the policymaker can obtain in a given state of the world if the regime shift has already occurred.

Prior to crossing a threshold, the policymaker distinguishes between the case where the system dynamics remain the same and the case where the system tips. The hazard rate $h(S_t, S_{t+1})$ gives the probability of tipping.\(^4\) By assumption, we have no record of events that previously triggered the tipping point, and abrupt regime shifts in system dynamics are hard to forecast accurately. The hazard rate of tipping points therefore exhibits deep uncertainty and the probabilities governing tipping points are not confidently known. Deep uncertainty is often associated with greater uncertainty aversion (Camerer and Weber 1992).

We use a normatively-founded rational model of deep uncertainty axiomatized in Traeger (2010), which relates closely to the smooth ambiguity model of Klibanoff et al. (2005, 2009).\(^5\) As Traeger (2010) explains, the model also relates closely to the Epstein

\(^4\)In general, the state space contains informational variables that tell the policymaker which part of the state space has already been explored. In our climate application, the policymaker keeps track of the greatest historic temperature.

\(^5\)The normative axiomatization relies on the von Neumann-Morgenstern axioms, but recognizes that
and Zin (1989) and Weil (1989) models that disentangle risk attitude from the policymaker’s propensity to smooth consumption over time. The long-run risk literature in asset pricing finds that risk aversion is significantly larger than the willingness to smooth consumption over time (e.g., Vissing-Jørgensen and Attanasio 2003, Bansal and Yaron 2004). In these models, a concave function \( f_{\text{unc}} \) captures how the policymaker is more averse to risk than to known changes in consumption over time. This function \( f_{\text{unc}} \) is even more concave for policymakers who are averse to deep uncertainty and face ambiguous hazards as in our setting. The main feature making our model one of deep uncertainty is that we apply the aversion function \( f_{\text{unc}} \) only to the subjectively uncertain tipping hazard, and not to the stochastic shocks \( \epsilon_t \). In the absence of the stochastic shocks, our results for uncertainty aversion could be interpreted as the effects of risk aversion when disentangled from willingness to substitute consumption over time.

In the pre-threshold regime, the policymaker solves the dynamic programming problem:

\[
V_0(S_t) = \max_{x_t} \left\{ u(x_t, S_t) + \beta \int f_{\text{unc}}^{-1}\left[ [1 - h(S_t, S_{t+1})] \ f_{\text{unc}}[V_0(S_{t+1})] \\
+ h(S_t, S_{t+1}) \ f_{\text{unc}}[V_1(S_{t+1})] \right] d\mathbb{P} \right\} \\
\text{s.t. } S_{t+1} = g_0(x_t, \epsilon_t, S_t) \\
\quad x_t \in \Gamma(S_t) .
\]

With probability \( h \) the system tips and \( V_1 \) determines the maximal expected future welfare from period \( t + 1 \) on. With probability \( 1 - h \), the system dynamics remain unaltered and \( V_0 \) characterizes future welfare. The expression under the integral evaluates the future as the non-linear weighted mean of the welfare derived in the case that the system stays unaltered and in the case that the system tips. The concave weighting function \( f_{\text{unc}} \) captures uncertainty aversion, as described above. The integral itself takes expectations over the stochastic shocks. In contrast to the distribution governing the chance of a regime shift, we assume that the distribution governing these temperature shocks is well understood, deriving from a long record of observations. These shocks are therefore not deeply uncertain or ambiguous, and their effects are evaluated outside the Knightian uncertainty (or ambiguity) operator \( f_{\text{unc}} \).

---

a decision-maker can have different confidence in different distributions.
2.2 The basic policy trade-off

We now identify the channels through which tipping points affect optimal policy. The next section explains how uncertainty aversion alters each channel. The possible existence of a tipping point introduces two new terms into the marginal welfare impact of changing a control. For ease of exposition, we analyze the case where a single state variable determines the chance of crossing the threshold. The right-hand side of equation (2) characterizes welfare for an optimal choice of the controls (with optimality denoted by \( \ast \)). We evaluate the marginal welfare impact of varying a generic entry \( e_t \) of the control vector in the neighborhood of the optimum. In our climate change application, the temperature state variable determines the hazard, and the welfare impact of varying emissions determines the optimal carbon tax. Suppressing all arguments independent of \( e_t \), the value of the optimal policy program is:

\[
\begin{align*}
&u(e^*_t) + \beta_t \int f^{-1}_{unc} \left[ 1 - h(S_{t+1}(e^*_t)) f_{unc}[V_0(S_{t+1}(e^*_t))] + h(S_{t+1}(e^*_t)) f_{unc}[V_1(S_{t+1}(e^*_t))] \right] d\mathbb{IP}. \\
&V_{eff}(e^*_t)
\end{align*}
\]

The integrand \( V_{eff}(e^*_t) \) expresses the value of future periods’ optimal policy program in utility units. It is the uncertainty-averse mean of pre- and post-threshold value. The total value aggregates \( V_{eff} \) over standard risk and combines it with the immediate utility from current choices. Varying \( e_t \) gives the following trade-off characterizing optimal policies:

\[
\frac{\partial u(e^*_t)}{\partial e_t} = -\beta \int \left\{ \left[ 1 - h(S_{t+1}(e^*_t)) \right] \frac{f'_{unc}[V_0(S_{t+1}(e^*_t))]}{f_{unc}[V_{eff}(e^*_t)]} \frac{\partial V_0(S_{t+1}(e^*_t))}{\partial S_{t+1}} \frac{\partial S_{t+1}(e^*_t)}{\partial e_t} \\
+ h(S_{t+1}(e^*_t)) \frac{f'_{unc}[V_1(S_{t+1}(e^*_t))]}{f_{unc}[V_{eff}(e^*_t)]} \frac{\partial V_1(S_{t+1}(e^*_t))}{\partial S_{t+1}} \frac{\partial S_{t+1}(e^*_t)}{\partial e_t} \\
- \frac{\partial h(S_{t+1}(e^*_t))}{\partial S_{t+1}} \frac{\partial S_{t+1}(e^*_t)}{\partial e_t} \frac{f_{unc}[V_0(S_{t+1}(e^*_t))] - f_{unc}[V_1(S_{t+1}(e^*_t))]}{f'_{unc}[V_{eff}(e^*_t)]} \right\} d\mathbb{IP},
\]

where primes (’) indicate derivatives.\(^6\) We interpret this equation for the case where an increase in \( e_t \) raises current utility but decreases expected future welfare. For instance, additional carbon dioxide emissions increase current utility but decrease future welfare by generating higher carbon stocks and temperatures; additional borrowing increases cur-

---

\(^6\)For a multidimensional state space, \( \partial V/\partial S_{t+1} \) and \( \partial h/\partial S_{t+1} \) denote gradients and \( \partial S_{t+1}/\partial e_t \) denotes the vector of state changes caused by the marginal change in \( e_t \). These derivatives are taken with respect to the pre-threshold dynamics \( g_0 \).
rent consumption but also increases future debt; and a political elite’s appropriation of resources and repression increase social discontent and hamper growth. By virtue of Bellman’s principle of optimality, equation (3) boils the complex intertemporal decision tree down into a two period trade-off. All future periods face a similar trade-off (conditional on not having tipped). These future trade-offs and policy responses are captured in the (pre- and post-threshold) value functions for period $t + 1$.

The left-hand side of equation (3) characterizes the (immediate) benefits from increasing the policy variable. At an optimum, these benefits must balance the expected future costs. The costs are represented by the right-hand side of equation (3) and are subject to uncertainty. The integrand in the first line represents the impact of policy on time $t + 1$ welfare under the pre-threshold regime (i.e., on the pre-threshold continuation value). This impact is composed of the control’s impact on the state vector and the effect of the altered state vector on pre-threshold value $V_0$, and it is weighted by the probability $(1 - h)$ of staying in the pre-threshold regime. In a world without tipping points (where the hazard rate $h$ is zero), the first line characterizes the full trade-off between current and future welfare.

Lemoine and Traeger (2014) explain that tipping points extend the standard trade-off in two ways. First, optimal policy now accounts for its ability to influence the endogenous probability of the regime shift. They call this channel the marginal hazard effect (MHE). It is captured in the third line of equation (3). When triggering a tipping point reduces welfare, this channel moves the optimal control in a direction that reduces the probability of tipping. Second, today’s policy affects conditions in the shifted regime in the case that tipping happens to occur. Lemoine and Traeger (2014) call this channel the differential welfare impact (DWI). It is captured by the terms with $h$ in the first two lines of equation (3). Its ambiguous sign depends on how changing the control affects welfare in the pre- and post-threshold worlds. We now show how uncertainty aversion changes these two tipping point contributions to optimal policy.

### 3 Uncertainty aversion and optimal policy in the face of a tipping point

This section studies the trade-off in equation (3) to show how uncertainty aversion affects optimal policy in the face of a possible tipping point. First, we explain how uncertainty

---

7In addition, the potential for future tipping points changes the pre-threshold continuation value and so also affects policy via the first line.
aversion affects optimal policy for an endogenous tipping hazard by changing the willingness to invest in reducing the probability of tipping. Second, we explain how uncertainty aversion affects optimal policy under a (possibly exogenous) tipping hazard by accounting for a policy’s impact on post-threshold welfare.

3.1 The policy implications of an endogenous hazard

When the probability of tipping risk is endogenous, the third line in equation (3) also affects policy: the optimal policy accounts for the marginal impact of the control \( e_t \) on the hazard rate. The following term reflects the endogenous hazard:

\[
MHE_{\text{total}} = - \frac{\partial h(S_{t+1}(e_t^*))}{\partial S_{t+1}} \frac{\partial S_{t+1}(e_t^*)}{\partial e_t} \frac{f_{\text{unc}}[V_0(S_{t+1}(e_t^*))] - f_{\text{unc}}[V_1(S_{t+1}(e_t^*))]}{f'_{\text{unc}}[V_{\text{eff}}(e_t^*)]} .
\]

(4)

Lemoine and Traeger (2014) derive the uncertainty-neutral form:

\[
MHE_{\text{neutral}} = \frac{\partial h(S_{t+1}(e_t^*))}{\partial S_{t+1}} \frac{\partial S_{t+1}(e_t^*)}{\partial e_t} [V_0 - V_1].
\]

They call it the marginal hazard effect (MHE). As they explain, the MHE is composed of the response of the hazard rate to a change in the state vector (term i), the response of the state vector to a change in the control (term ii), and the total welfare change from switching regimes (term iii). The uncertainty-neutral marginal hazard effect always strengthens the optimal policy if raising the control increases the chance of triggering a harmful tipping point.

We analyze how uncertainty aversion affects this marginal hazard effect via term iii in equation (4):

\[
\frac{f_{\text{unc}}[V_0] - f_{\text{unc}}[V_1]}{f'_{\text{unc}}[V_{\text{eff}}]} = [V_0 - V_1] \frac{f_{\text{unc}}[V_0] - f_{\text{unc}}[V_1]}{[V_0 - V_1] f'_{\text{unc}}[V_{\text{eff}}]} .
\]

The fraction characterizes the contribution of uncertainty aversion. It is unity for an uncertainty-neutral decision-maker and when \( f'_{\text{unc}}[V_{\text{eff}}] = \frac{f_{\text{unc}}[V_1] - f_{\text{unc}}[V_0]}{|V_1 - V_0|} \). Given the concavity of \( f_{\text{unc}} \), this condition has to be satisfied for some \( V_{\text{eff}} \) and, thus, for some hazard \( h^{\text{critical}} \). If the hazard is lower than this critical value, then \( V_{\text{eff}} \) increases and the denominator \( f'_{\text{unc}}[V_{\text{eff}}] \) falls. Thus, for hazard rates below the critical value, uncertainty aversion
amplifies the marginal hazard effect. For hazards above the critical level, uncertainty aversion reduces the marginal hazard effect.

Using a second-order Taylor expansion of $f_{\text{unc}}$ around $V_{\text{eff}}$, we approximate the uncertainty multiplier as

$$
\frac{f_{\text{unc}}[V_0] - f_{\text{unc}}[V_1]}{V_0 - V_1} f_{\text{unc}}'[V_{\text{eff}}] \approx 1 + \left| \frac{f_{\text{unc}}''}{f_{\text{unc}}'} \right|_{V_{\text{eff}}} \left( V_{\text{eff}} - \frac{V_0 + V_1}{2} \right)
$$

and obtain the overall impact from the marginal hazard effect

$$
MHE_{\text{total}} \approx MHE_{\text{neutral}} \left[ 1 + MHE_{\text{unc}} \right].
$$

The new term $MHE_{\text{unc}}$ is positive for hazards smaller than a critical value which is less than one half: the generalized mean $V_{\text{eff}}$ places only the weight $h$ on the bad outcome as compared to the arithmetic mean $\frac{V_0 + V_1}{2}$ that places equal weights on the two states of the world. Increasing uncertainty aversion increases the uncertainty contribution through the measure of absolute uncertainty aversion $\left| \frac{f_{\text{unc}}''}{f_{\text{unc}}'} \right|_{V_{\text{eff}}}$. However, it also reduces the contribution of the term in brackets by reducing the uncertainty-averse mean $V_{\text{eff}}$. As a consequence, the marginal hazard effect is nonconvex in uncertainty aversion.

The uncertainty multiplier has a similar intuition as the expression for the willingness to pay for a risk reduction in the standard risk setting. A risk-averse agent is not always willing to pay more for a risk reduction than is a risk-neutral agent. The more risk-averse (uncertainty-averse) the decision-maker, the more he values wealth in bad states relative to good states. Giving up wealth for a hazard reduction makes a decision-maker even worse off in the bad state of the world. If a hazard is large enough, he prefers carrying wealth into the post-threshold world over spending it on reducing the hazard. This effect of uncertainty aversion on the MHE has parallels in the literature studying self-protection under uncertainty aversion (Treich 2010, Snow 2011, Alary et al. 2013, Maccheroni et al. 2013). In these settings, uncertainty aversion skews optimal policy towards reducing uncertainty. We see the same type of effect here.

### 3.2 The policy implications of the post-threshold world

In a world without uncertainty aversion, a potential tipping point also changes the policy trade-off through the terms proportional to the hazard rate in the first two lines of
equation (3):

\[ DWI_{\text{neutral}} = h \left[ \left( -\frac{\partial V_1}{\partial e_t} \right) - \left( -\frac{\partial V_0}{\partial e_t} \right) \right] = h \left[ \frac{\partial V_0}{\partial S_{t+1}} - \frac{\partial V_1}{\partial S_{t+1}} \right] \frac{\partial S_{t+1} (e_t^*)}{\partial e_t}. \] (5)

Lemoine and Traeger (2014) call this adjustment the differential welfare impact (DWI) because it is proportional to the difference in the marginal impact of the control on the pre- and post-threshold value functions. This differential welfare impact strengthens policy if relaxing the control has a more harmful impact in the the post-threshold regime. Note that this term does not depend on whether the probability of tipping is exogenous or endogenous.

We now analyze how uncertainty aversion changes the differential welfare impact. An uncertainty-averse policymaker is additionally averse to the types of poorly understood uncertainty that characterize tipping points. A strictly concave function \( f_{\text{unc}} \) captures her additional aversion to tipping point uncertainty. Our extended definition of DWI now collects all the changes in the first two lines of equation (3) with respect to the continuation value:

\[ DWI_{\text{total}} = DWI_{\text{neutral}} + h \left[ \frac{f_{\text{unc}}' (V_1)}{f_{\text{unc}}' (V_{eff})} - 1 \right] \left( -\frac{\partial V_1}{\partial e_t} \right) + (1 - h) \left[ \frac{f_{\text{unc}}' (V_0)}{f_{\text{unc}}' (V_{eff})} - 1 \right] \left( -\frac{\partial V_0}{\partial e_t} \right). \] (6)

These changes capture the effect of an exogenous hazard and its interaction with uncertainty aversion. For an uncertainty-averse decision-maker who faces a welfare-decreasing tipping point, the concavity of \( f_{\text{unc}} \) implies \( f'(V_0) < f'(V_{eff}) < f'(V_1) \). Therefore, the first term’s contribution to \( DWI_{\text{unc}} \) in equation (6) is positive, whereas the second term’s contribution is negative.

We now identify situations where \( DWI_{\text{unc}} \) is negative. We assume that uncertainty aversion does not fall too quickly in welfare and that the tipping point does not cause too large a welfare loss. Then the following approximation characterizes the uncertainty
contribution:\(^8\)

\[
DWI_{\text{unc}} \approx \frac{-f''_{\text{unc}}}{f'_{\text{unc}}} \bigg|_{V_{\text{eff}}} \left[ \frac{h(V_{\text{eff}} - V_1)}{w^-} \left( -\frac{\partial V_1}{\partial e_t} \right) - \frac{(1 - h)(V_0 - V_{\text{eff}})}{w^+} \left( -\frac{\partial V_0}{\partial e_t} \right) \right]. \tag{7}
\]

This contribution is proportional to the measure of absolute uncertainty aversion \(-\frac{f''_{\text{unc}}}{f'_{\text{unc}}}\) evaluated at \(V_{\text{eff}}\). The terms in square brackets decide the sign of the contribution. Similarly to the uncertainty-neutral contribution \(DWI_{\text{neutral}}\), this sign depends on the relative magnitudes of the control’s marginal welfare impact in the pre- and post-threshold worlds. However, in contrast to the uncertainty-neutral contribution, these marginal welfare impacts are weighted by the terms \(w^-\) and \(w^+\), respectively.

Under uncertainty neutrality, we have \(w^- = w^+ = h(1 - h)(V_0 - V_1)\). Hence, for an uncertainty-averse decision-maker it is \(w^+ > w^-\) because \(V_{\text{eff}}\) is smaller than the arithmetic mean. As a consequence, the uncertainty-averse contribution favors the marginal pre-threshold impact when determining the overall sign of the contribution. If this pre-threshold welfare impact of the control is larger than its post-threshold impact, then \(DWI_{\text{neutral}} < 0\). Thus, equation (7) implies that, if the uncertainty-neutral contribution to the differential welfare impact is negative, then the uncertainty-averse contribution is also negative (\(DWI_{\text{unc}} < 0\)).

In general, the sign of the uncertainty-neutral contribution \(DWI_{\text{neutral}}\) is ambiguous. In our climate change example, a tipping point that makes carbon dioxide more harmful in the shifted regime would favor an increase in the carbon tax. At the same time, the endogenous tipping hazard makes a carbon dioxide reduction in (and only in) the pre-threshold regime more valuable because of the reduction of future\(^9\) tipping risks. If the marginal effect of emissions on the tipping hazard is sufficiently feared, then the differential welfare impact’s adjustment reduces the optimal carbon tax: the DWI accounts for the possibility that the policymaker crosses into the post-threshold scenario and no longer benefits from a policy induced hazard reduction. The uncertainty averse contribution shifts slightly earlier than the uncertainty neutral contribution as uncertainty aversion increases the policymaker’s fear of tipping (represented through ambiguity aversion and the weights \(w^-\) and \(w^+\)).

---

\(^8\)We approximate \(f'_{\text{unc}}(V_i)\) for \(i \in \{0, 1\}\) using the measure of absolute uncertainty aversion \(-\frac{f''_{\text{unc}}}{f'_{\text{unc}}}\) evaluated at \(V_{\text{eff}}\) to obtain \(\frac{f'_{\text{unc}}(V_i)}{f_{\text{unc}}(V_{\text{eff}})} \approx \frac{f''_{\text{unc}}}{f'_{\text{unc}}}(V_i - V_{\text{eff}}), i \in \{0, 1\}\).

\(^9\)The MHE captures the immediate tipping risk in a given period. However, a ton of carbon dioxide released in the current period also implies a tipping risk in future periods. The pre-threshold value function’s response to an increase in the carbon dioxide captures this future tipping risk.
In our climate change application, the DWI increases the optimal carbon tax under uncertainty neutrality. The effect is driven by the stronger harm from emissions in the post-tipping world. Under uncertainty aversion, the DWI's sign switches because the policymaker becomes more afraid of tipping. This sign switch results from an increase in the pre-threshold value function's sensitivity to the tipping hazard. The MHE remains the primary channel through which tipping points affect the optimal tax on carbon dioxide emissions. It increases the cost of emissions because of the endogenous tipping hazard in the pre-threshold world, which the uncertainty-averse policymaker especially fears. If the system tips (at hazard rate $h$) the policymaker no longer benefits from a policy-induced reduction in the tipping hazard.

The effect of uncertainty aversion on the DWI has parallels in the literature studying self-insurance under uncertainty aversion (Gollier 2011, Traeger 2011, Alary et al. 2013). Our setting differs by considering the full intertemporal interaction between self-protection and self-insurance in a world with an endogenous tipping hazard. In particular, the cited literature generally finds that self-insurance increases in uncertainty aversion. In our setting the DWI increases only if the policy has a strong enough marginal welfare impact in the post-threshold world to dominate the fear of the endogenous tipping hazard. This condition is not satisfied in our climate change application.

4 A climate-economy model with tipping points and uncertainty aversion

We now consider the effect of climate tipping points on the optimal tax on carbon dioxide emissions. The optimal carbon tax equals the social cost of carbon when evaluated along the optimal policy path in a welfare-maximizing integrated assessment model. We here only summarize the model structure, referring the reader to Lemoine and Traeger (2014) for the complete mathematical description and computational methods.

We reformulate the benchmark Dynamic Integrated model of Climate and the Economy (DICE) from Nordhaus (2008) as an infinite-horizon dynamic programming problem with a tipping point in the climate system, optimal learning about the threshold that triggers a tipping point, and a generalized welfare evaluation. DICE is a Ramsey-Cass-Koopmans growth model that has an aggregate world economy interacting with a climate module. In each period, a policymaker allocates available output to consumption, investment in a durable capital stock, and to reducing emissions of carbon dioxide (CO$_2$). Emissions increase the stock of CO$_2$, which generates warming in future periods. Warm-
ing in turn reduces output. Reducing emissions therefore acts like investment in that it transfers wealth from the present to the future. The pre-threshold world has standard DICE dynamics along with the tipping possibility and temperature shocks calibrated to the historical record.

A tipping point irreversibly changes the climate system from its conventional representation in DICE to a new regime with altered dynamics. The tipping point occurs upon crossing some unknown temperature threshold. Current emissions therefore increase the chance of triggering a tipping point in some future period. The timing, probability, and welfare consequences of a regime switch are endogenous because they depend on the policies chosen before and after the threshold occurs. We evaluate two tipping points of concern in the climate science literature. In every model run, the policymaker faces a single tipping point and knows in advance what its effects would be. Lemoine and Traeger (2014) motivate each of the tipping points with reference to the scientific literature and describe the calibration of the endogenous hazard rate. They explore three strengths for each type of tipping point and discuss the implications of alternative modeling assumptions.

The first tipping point increases the climate feedbacks that amplify global warming, and the second increases the atmospheric lifetime of CO\textsubscript{2}. The first tipping point therefore increases the effect of emissions on temperature, and the second increases the time during which emissions affect the climate. Temperature dynamics in DICE depend on a parameter known as climate sensitivity, which is the equilibrium warming from doubling CO\textsubscript{2}. We represent a climate feedback tipping point as increasing climate sensitivity from its pre-tipping value of 3°C to either 4°C or 5°C. The second tipping point reflects the possibility that carbon sinks weaken beyond the predictions of coupled climate-carbon cycle models, which is similar to decreasing the “decay rate” of CO\textsubscript{2}. We represent these weakened sinks by decreasing the transfer of CO\textsubscript{2} out of the atmosphere by 25%, 50%, or 75%.

As noted in discussing equation (2), the function \( f_{\text{unc}} \) captures aversion to the Knightian uncertainty surrounding tipping points, which are less understood than other climate phenomena (Alley et al. 2003, Lenton et al. 2008, Ramanathan and Feng 2008, Kriegler et al. 2009, Smith et al. 2009). Sticking to isoelastic preferences, we adopt the power function \( f_{\text{unc}}(V) = ((1 - \eta)V)^{1-\eta}. \) Here \( \gamma \) is a measure of Arrow-Pratt relative risk aversion with respect to Knightian uncertainty, and \( \eta = 2 \) is the constant Arrow-Pratt measure of relative risk aversion used in the standard DICE utility function. The Knightian uncertainty surrounding tipping points contrasts with the standard risk posed by the
historically-grounded temperature stochasticity and governed by \( \mathbb{P} \). If Arrow-Pratt risk aversion with respect to subjective risk \( \gamma \) coincides with standard risk aversion \( \eta \), then the function \( f_{\text{unc}} \) is linear and drops out. In that case, the policymaker is uncertainty-neutral and the welfare evaluation is as in DICE. However, when \( \gamma > \eta \), the function \( f_{\text{unc}} \) measures the policymaker’s additional aversion to uncertainty (or subjectivity of belief) governing tipping points as opposed to annual temperature stochasticity.

5 The optimal carbon tax when facing possible tipping points

We compare several sets of model runs to assess how the optimal carbon tax responds to aversion to tipping point uncertainty. The optimal carbon tax is the policy trajectory that maximizes the present value of net benefits within our extension of the benchmark DICE integrated assessment model. All of our graphs present results conditional on not having crossed the threshold: we want to understand how optimal policy changes in the face of a potential tipping point. The depicted paths draw the multiplicative temperature shock at its expected value in each period.

Figure 1 gives the effect of tipping points and uncertainty attitudes on the optimal carbon tax (social cost of carbon), the optimal \( \text{CO}_2 \) concentration path, and the optimal temperature path. All trajectories use the “middle” strength version of each tipping point. These plots vary \( \gamma \) from uncertainty-neutral (\( \gamma = 2 = \eta \)) to extremely uncertainty-averse (\( \gamma = 100 > 2 = \eta \)). We find that optimal policy is not highly sensitive to the policymaker’s level of uncertainty aversion. The near-term carbon tax varies by less than $1/\text{tCO}_2$ across the modeled range of \( \gamma \), the peak \( \text{CO}_2 \) concentration varies by less than 30 ppm, and peak temperature varies by less than 0.2°C. For the climate feedback tipping point, the extreme form of uncertainty aversion with \( \gamma = 100 \) increases the optimal carbon tax by 6% in 2015 and by 12% in 2050. For the carbon sink tipping point, the extreme form of uncertainty aversion also increases the optimal carbon tax by around 6.5% in both 2015 and 2050. The effect on near-term policy of varying uncertainty aversion across our scenarios is slightly smaller than the effect of varying the strength of a tipping point in Lemoine and Traeger (2014).

Figure 2 plots the year 2015 optimal carbon tax and the peak temperature for values of \( \gamma \) between 0 (uncertainty-loving) and 100 (strongly uncertainty-averse), with all calculations conditional on not having crossed the threshold. This figure depicts the results for three different strengths of the carbon sink tipping point and for two different strengths.
Figure 1: Time paths for the optimal carbon tax (current value), the CO₂ stock, and temperature under each type of tipping point possibility, using expected draws. We simulate a path that happens to never cross a threshold in order to see how the modeled policymaker adjusts to the possibility over time. The parameter \( \gamma \) controls the strength of uncertainty aversion, ranging from uncertainty neutrality (\( \gamma = 2 \)) to extreme uncertainty aversion (\( \gamma = 100 \)).
Figure 2: The optimal carbon tax in 2015 and the peak temperature reached for each type of uncertainty attitude. The plotted simulations assume expected draws of the temperature shock and assume that the tipping point never occurs. The parameter $\gamma$ controls the strength of uncertainty aversion, ranging from weakly uncertainty-loving ($\gamma = 0$) to uncertainty-neutral ($\gamma = 2$) to extremely uncertainty-averse ($\gamma = 100$).

As the level of uncertainty aversion increases, the optimal year 2015 carbon tax becomes slightly stronger. The greatest effects arise in the case of the strongest carbon sink tipping point, with the plotted cases showing that the optimal tax increases from $14/\text{tCO}_2$ when $\gamma = 0$ to over $16/\text{tCO}_2$ when $\gamma = 100$. We also see that greater uncertainty aversion decreases optimal peak temperature slightly in all cases. The greatest effects on peak temperature arise for stronger versions of the feedback tipping point, with the plotted case showing that peak temperature falls from 2.9°C in the case of $\gamma = 0$ to 2.7°C in the case of $\gamma = 100$.

We now use our decomposition of the optimal emission tax from Section 3 to assess the channels through which uncertainty aversion affects policy (Figure 3). Each panel depicts either the differential welfare impact (DWI) or the marginal hazard effect (MHE) in the

---

When the feedback tipping point increases climate sensitivity to 6°C, the large welfare losses associated with tipping cause computational problems in the presence of a high degree of uncertainty aversion.
case of uncertainty neutrality ($\gamma = 2$) and in the case of extreme uncertainty aversion ($\gamma = 100$). Begin with the MHE (top row). In keeping with our analysis, we see that the total MHE is positive in all cases and that uncertainty aversion can substantially increase the MHE. Further, the MHE rises strongly over time as the welfare loss from tipping increases and as the marginal effect of emissions on the hazard increases as a consequence of learning about the threshold’s location (see Lemoine and Traeger 2014). The MHE grows faster in the face of the feedback tipping point than in the face of the carbon sink tipping point because the welfare loss imposed by the feedback tipping point increases strongly in CO$_2$. Once the tipping point has occurred, high CO$_2$ concentrations cause much higher damages than in the pre-threshold regime. In contrast, the welfare loss from the carbon sink tipping point starts out relatively high, but it does not grow as fast over time because the policymaker’s ability to control the cost of tipping is not as state-dependent.$^{11}$

Now consider the DWI (middle row). It is typically more than an order of magnitude smaller than the MHE. In the case of uncertainty neutrality, the DWI is mostly positive because emission reductions pay off especially strongly if a tipping point happens to increase warming per unit of emissions or decrease the effectiveness of carbon sinks. The exception arises for the carbon sink tipping point late in the century. There, we see a slightly negative DWI because the tipping point makes carbon dioxide reach very high levels, which reduces the marginal effect of another unit of carbon dioxide on temperature.$^{12}$

Under strong uncertainty aversion, the total DWI reduces the optimal carbon tax, through both its uncertainty-averse and uncertainty-neutral components. The discussion in section 3.2 explain such a sign switch of the DWI under uncertainty aversion. The pre-threshold value function becomes more sensitive to the control than the post-threshold value function because the policymaker is more strongly affected by the tipping hazard. Uncertainty aversion amplifies the policymaker’s fear of future tipping, a hazard that

$^{11}$See the discussion in Lemoine and Traeger (2014) for more on how the two types of tipping points affect policy in qualitatively different ways.

$^{12}$Additional units of carbon dioxide generate warming by trapping outgoing infrared radiation (heat). when there is already a lot of CO$_2$ in the atmosphere, the stock of CO$_2$ already nearly completely blocks the wavelengths over which CO$_2$ most effectively traps heat. Additional emissions therefore primarily contribute to warming by blocking wavelengths at which they are less effective, so that the marginal contribution of emissions to warming declines in the stock of CO$_2$. The negative DWI tells us that the marginal impact of emissions in the post-threshold world no longer dominates its impact in the pre-threshold world, which includes both the immediate impact on welfare in the pre-threshold world as well as the reduction of future tipping risk that is also captured in the pre-threshold value function’s response to emissions.
Figure 3: Time paths for MHE (top) and DWI (middle) for tipping points that increase climate sensitivity to 5°C (left) and that weaken carbon sinks by 50% (right). The bottom row shows DWI calculated using a “no-threshold” continuation value for the pre-tipping regime, as discussed in the text. The lines labeled “Unc” depict the contribution of uncertainty aversion in cases with $\gamma = 100$. Note that the scales of the vertical axes do not match each other.
remains as long as she is still in the pre-tipping world (and temperatures are increasing). This impact is measured by the MHE. The differential welfare impact corrects for the fact that, with probability \( h \), the policymaker will no longer be in this pre-threshold world. In this post-threshold world, abatement no longer provides the additional payoff of reducing the tipping hazard. If the (pre-threshold) fear of tipping is sufficiently large, then it dominates the severity of post-threshold carbon dioxide emissions and accounting for the differential welfare impact reduces the optimal carbon tax (\( DWI < 0 \)).

The bottom row of Figure 3 demonstrates the underlying reasoning. Instead of using the pre-threshold continuation value \( V_0(S_{t+1}) \), it calculates a “DWI equivalent” using the continuation value that would obtain if future tipping was not possible and the only hazard would be to tip in the current period. This calculation removes the future tipping risk from the pre-threshold value function. In this hypothetical world, both the uncertainty-neutral and the uncertainty-averse DWI contributions turn positive again: the impact of a ton of carbon dioxide is strongest in the post-threshold world.

6 Conclusions

We have shown how aversion to uncertainty (or ambiguity) affects optimal policy in the face of poorly understood tipping points. First, uncertainty aversion affects optimal policy through the marginal hazard effect (MHE) when the probability of triggering a tipping point is endogenous. Uncertainty aversion tends to move policy in a direction that reduces Knightian uncertainty. More precisely, for a tipping hazard that is not too large, uncertainty aversion pushes the policymaker to spend more on reducing the endogenous hazard. If the hazard becomes large enough, the value of transferring consumption into the bad state of the world can outweigh the self-protection incentive and reduce the hazard reduction effort. This effect increases in the sensitivity of the tipping hazard to the policy control, in the expected welfare loss from tipping, and, at least for small hazard rates, in uncertainty aversion. These findings relate closely to the usual arguments about self-protection under risk and ambiguity.

Second, uncertainty aversion changes the differential welfare impact (DWI) of the optimal policy, which accounts for the policy’s post-threshold value. If the payoffs from a stringent policy are higher after the regime shift than in the pre-threshold regime, then the DWI strengthens the optimal policy if the chance of future tipping is not too high and the policymaker is not too uncertainty-averse. However, strong aversion to an endogenous tipping hazard can flip this sign. The fear of eventual tipping makes a strong
policy particularly valuable in the pre-threshold (and only in the pre-threshold) world. Then the post-threshold impact adjustment captured by the DWI reduces the strength of the optimal policy.

In our numeric climate change application, we find that uncertainty aversion slightly increases the optimal carbon tax. The uncertainty aversion’s contribution to the MHE always dominates in magnitude and strengthens the policy. The DWI is slightly positive under uncertainty neutrality, but it reduces the optimal carbon tax in the case of an uncertainty-averse policymaker. The lack of knowledge governing tipping point locations in the climate system is severe. We find that a framework that explicitly incorporates the Knightian nature of this uncertainty leads to only small (upwards) adjustments of the optimal carbon tax. Jensen and Traeger (2014) find similarly small adjustments in a world of smooth climate change, and we show that these adjustments remain small even when there is a significant hazard of major abrupt losses induced by a regime shift.

Our result suggests that modelers and policymakers should not shy away from following a best-guess Bayesian prior approach to tipping points to derive and follow optimal mitigation policy. The result does not suggest that better information on the nature of the tipping points or their location is not valuable. We evaluated the policy impact of Knightian uncertainty, which is moderate. Lemoine and Traeger (2014) show that a better estimate of the Bayesian prior can have a stronger impact on the optimal policy than the adjustments we obtain here from accounting for or eliminating the Knightian nature of the uncertainty.

References


Bansal, Ravi and Amir Yaron (2004) “Risks for the long run: A potential resolution of


Lange, Andreas and Nicolas Treich (2008) “Uncertainty, learning and ambiguity in eco-


