Public Goods in the Short and Long Term



Philip Trammell
New College
University of Oxford

Submitted for the degree of
Doctor of Philosophy in Economics
October 2024

Supervisors

This thesis was supervised by Ian Jewitt and Alex Teytelboym.

Statement of authorship

The thesis consists of three papers. The first, "Existential Risk and Growth", is coauthored with Leopold Aschenbrenner. The second and third are single-authored.

Statement of previous acceptance

An earlier draft of the second paper, "Dynamic Public Good Provision under Time Preference Heterogeneity", was accepted as a thesis for the MPhil in Economics, University of Oxford, 2021.

Funding

I gratefully acknowledge funding for this degree from the Atkinson Scholarship (Global Priorities Institute, University of Oxford) and the Early Career Funding Program (Open Philanthropy).

Acknowledgments

I greatly thank William MacAskill, for encouraging me to move to Oxford and to work on questions like those addressed in this thesis; many other researchers—senior and junior, academic and applied—who have passed through Littlegate House and Trajan House, for inspiring almost all my research ideas (and all my best ones) and giving feedback on my drafts; my supervisors, for extensive advice and support despite my relatively idiosyncratic range of projects; and the family and friends who have helped me distinguish the times anxiety is misleading me from the rarer times there really is something to worry about.

Contents

Thesis abstract **1**

I Existential Risk and Growth **2**

1 Introduction **3**

2 State risk without policy **7**
2.1 Model 7
2.2 How does acceleration affect risk? 8

3 State risk with policy **12**
3.1 Motivation 12
3.2 The economic environment 14
3.3 The existential risk Kuznets curve 17
3.4 Acceleration and state risk 24
3.5 Discussion 30

4 Transition risk **31**
4.1 Motivation 31
4.2 A transition-risk-based hazard function 31
4.3 Acceleration and transition risk 33
4.4 Discussion 37

5 Conclusion **38**

A Supplemental materials **40**
A.1 Calibrating the elasticity of the hazard rate to safety expenditures . . 40
A.2 State risk with policy: growth vs. patience 42
A.3 State risk with policy: generalized results 43
A.4 State risk with policy: Two more hazard functions of interest 53
A.5 Transition risk: Optimal technology growth 59

B Proofs **66**
B.1 Existence and uniqueness of optimal policy path 66
B.2 Proof of Proposition 1 70
B.3 Proof of Proposition 3 73
B.4 Proof of Proposition 4 75
B.5 Proof of Proposition 5a 77
B.6 Proof of Proposition 6 80

B.7 Proof of Proposition 7 . . . 84
B.8 Safety in redundancy . . . 87

C Transition dynamics for simulations **90**

II Dynamic Public Good Provision under Time Preference Heterogeneity 93

1 Introduction **94**
1.1 Motivation . . . 94
1.2 Related discounting literature . . . 98
1.3 Outline . . . 101

2 Benchmarks **102**
2.1 Private goods . . . 104
2.2 Open-loop game: intertemporal polarization . . . 108
2.3 Stackelberg game: first-mover advantage . . . 111

3 Dynamic game **113**
3.1 Setup . . . 113
3.2 The polarized equilibrium . . . 123
3.3 Superior equilibrium payoffs . . . 124

4 Constraints **126**
4.1 Spending maxima / borrowing constraints . . . 128
4.2 Spending minima / saving constraints . . . 134
4.3 The costs of constraints . . . 137

5 Delayed benefits **142**

6 Conclusion **146**

A Proofs **150**
A.1 Proof of Proposition 1 and corresponding payoff . . . 150
A.2 Proof of Proposition 2 . . . 152
A.3 Proof of Proposition 3 . . . 153
A.4 Proof of Proposition 4 . . . 156
A.5 Proof of Proposition 5 . . . 179
A.6 Proof of Proposition 6 . . . 223
A.7 Proof of Proposition 7 . . . 240
A.8 Proof of Proposition 8 . . . 245
A.9 Proof of Proposition 9 . . . 248

A.10 Proof of Proposition 10 . . . 249
A.11 Proof of Proposition 11 . . . 261
A.12 Proof of Proposition 12 . . . 262

III Ethical Consumerism 264

1 Introduction 265
1.1 Motivation . . . 265
1.2 Related literature . . . 268
1.3 Outline . . . 273

2 Model 274
2.1 Economic environment . . . 274
2.2 Strategic demand function choice . . . 281
2.3 Summary . . . 287

3 Equilibrium 289
3.1 Characterization . . . 289
3.2 Refinement . . . 296
3.3 Existence . . . 301

4 Implications for consumer behavior 307
4.1 Prices and external preferences . . . 308
4.2 Anything goes . . . 310
4.3 Satiation . . . 312

5 Conclusion 318

A Supplemental materials 322
A.1 The normal form game . . . 322
A.2 Externality weights in elasticity terms . . . 325

B Proofs 327
B.1 Proof of Proposition 1 . . . 327
B.2 Proof of Proposition 2 . . . 328
B.3 Proof of Proposition 3 . . . 330
B.4 Proof of Proposition 5 . . . 337
B.5 Proof of Proposition 6 . . . 354
B.6 Proof of Proposition 9 . . . 362

References 370

Thesis abstract

This thesis consists of three papers relevant to the private provision of public goods.

The first considers how providing one public good (technological development) can realize another (existential safety: roughly, avoidance of a catastrophe that kills everyone). Technology may directly increase existential risk, but, by enriching society and increasing a planner's willingness to pay for safety, indirectly reduce existential risk. The second effect is found to outweigh the first surprisingly often.

The second paper studies a game in which funders with different discount rates provide a public good over time. Impatient funders should often crowd patient funders out of contributing in the short term, with the latter investing all their resources for future spending. Existing and proposed laws preventing the patient from doing this are almost as costly to them as full expropriation, if they are small.

The third extends a static public good game to a general equilibrium setting in which prices depend on expenditures. Accounting for this endogeneity motivates spending behavior markedly different both from equilibrium behavior in classic or partial equilibrium public good models, and from that in the Walrasian setting of general equilibrium with only private goods. Equilibrium is compatible with satiation and with philanthropically-motivated spending on public bads.

All three are theory papers, but little else unites them methodologically. Their common thread is a motivation: to improve how classical utilitarians spend their resources, given the complex relationships spending in one time and place can have to other markets or other times. They are thus lighter on policy prescriptions and welfare analyses than research on public good games often is. Their primary intended audience is not a planner seeking the efficient frontier, but a philanthropist acting on particular preferences in a partially non-cooperative setting.

I Existential Risk and Growth

Abstract

Technologies can pose existential risks to civilization. Though accelerating technological development may increase the risk of anthropogenic existential catastrophe per period in the short run, two considerations suggest that a sector-neutral acceleration decreases the risk that such a catastrophe *ever* occurs. First, acceleration decreases the time spent at each technology level. Second, since a richer society is willing to sacrifice more for safety, optimal policy can yield an "existential risk Kuznets curve"; acceleration then pulls forward a future in which risk is low. Acceleration typically increases risk only given sufficiently extreme policy failures or direct contributions of acceleration to risk.

Leopold Aschenbrenner and I thank the Centre for Effective Altruism, the University of Oxford's Future of Humanity Institute, and Columbia ISERP for their support. We also thank Danny Bressler, Lennart Stern, and Michael Wiebe for suggesting the idea; many people at the Global Priorities Institute, Ben Snodin, Luis Mota, Tyler Cowen, Rick van der Ploeg, Pete Klenow, and especially Chad Jones for helpful comments; and Alex Holness-Tofts and Arvo Muñoz for research assistance on an earlier draft.

1 Introduction

Technology brings prosperity. Its impact on *existential risk*—the risk of human extinction, or, equivalently for decision purposes, of an equally complete and permanent loss of human welfare—strikes many as ambiguous at best.[1] Advances in vaccine technology render us less vulnerable to devastating plagues; advances in gain-of-function virology arguably make them more likely (Millett and Snyder-Beattie, 2017).

This raises the possibility of a tradeoff: concern for the survival of civilization may motivate slowing development, at least outside some narrow domains. Environmentalist sentiments along these lines go back at least to the Club of Rome's 1972 report on the "Limits to Growth", and have arguably reemerged with calls to pause AI development (Future of Life Institute, 2023). Jones (2024) explores how to trade off between AI development and AI risk, assuming the tradeoff exists.

Does it? When would slowing technological development reduce existential risk?

To shed light on this question, we begin in Section 2 with a simple model in which the hazard rate—the probability of catastrophe per period—is a positive function of the technology level. Here, an existential catastrophe must occur eventually unless in the long run higher technology levels carry hazard rates that fall toward zero.

In this setting there are only two possibilities. If advanced technology does not eventually drive the hazard rate toward zero, then a catastrophe is inevitable, so accelerating technological development cannot increase its probability. Otherwise, a

[1] See e.g. Bostrom (2002), Posner (2004), Farquhar et al. (2017), Ord (2020), and Jones (2024). We will refer to the event that humanity immediately goes extinct or suffers a similarly complete and permanent loss of welfare as an "existential catastrophe", or simply "catastrophe". We will refer to "humanity" and "human civilization" interchangeably. We will not consider possible sources of value beyond human welfare, such as the welfare of other species or any intrinsic value placed on the preservation of the environment.

catastrophe is avoidable, and acceleration can lower its probability by hastening the arrival of safety.

This simple model formalizes two observations. First, if we believe the hazard rate is currently high, our only hope for a long future is the hope that we are in a temporary "time of perils". This view was famously expressed by Sagan (1997), and its implications for those especially concerned about the long-term future are emphasized by Parfit (1984), Ord (2020), and others. The second is less widely appreciated: that if we are in a time of perils, with the hazard rate a positive function only of the technology level, then deceleration for the sake of long-term survival is misguided. Speeding technological development may be temporarily risky, but in this setting it must be safer in the long run.[2]

The model of Section 2 is not "economic". It studies the impact on risk of quickly escaping risky states, not optimal policy under constraints. It thus leaves open the possibility that, when consumption–risk tradeoffs are navigated by a planner with little concern for long-term survival, technological acceleration can increase risk after all. Section 2 also offers no reason to believe that future states *will* be safe. If one believes that technology has historically increased the hazard rate, the hope that this relationship will reverse in the future may seem naive.

In Section 3, therefore, we introduce an environment in which technology grows exogenously and its risks can be mitigated by policy. As dangerous new technologies are introduced, a planner, discounting the future at an arbitrary rate, decides how much consumption to sacrifice to lower the hazard rate.

We illustrate that, even if technological advances in isolation always raise the hazard rate, optimal policy can generate an "existential risk Kuznets curve", with

[2]The point is however noted informally by Bostrom (2014), p. 234, and recently by Ord (2024).

the hazard rate rising and then falling as technology advances. Early, when the expected discounted value of the future of civilization is low and the marginal utility of consumption is high, it is worthwhile to adopt risky technologies as they arrive. Later, when the discounted future is more valuable and the marginal utility of consumption has fallen, substantial risk mitigation becomes worthwhile.

The possibility of rational policy thus offers an economic justification for the view that we may indeed be living through a once-in-history time of perils. Safety is a luxury. Technological development generates a wealth effect. If the utility function is concave enough—i.e. the wealth effect strong enough—optimal policy lowers the hazard rate quickly enough that the probability of long-term survival is positive.

This insight mirrors the logic of Stokey (1998) and Brock and Taylor (2005), on which environmental damages rise and then fall with economic development, and of Jones (2016, 2024), on which growth increases the value of life relative to marginal consumption. Like the analysis presented here, these papers find that, given a concave enough utility function, enrichment motivates large reallocations from consumption to safety. None of these sources solve for the optimal path of a hazard rate over time, however, or characterize conditions under which the probability of a binary event (here, existential catastrophe) under optimal policy is less than 1.

Our model of catastrophic risk differs more significantly from those of Martin and Pindyck (2015, 2021) and Aurland-Bredesen (2019). That literature studies a society's willingness to pay to reduce the risk of catastrophes that are, or are equivalent to, proportional consumption cuts. In such a context there are no wealth effects: the fraction of consumption one is willing to sacrifice to avoid a proportional consumption cut is, by definition, independent of one's consumption.

For the reasons that policy facilitates survival on a *given* (increasing) technology

path, optimal policy tends to magnify the extent to which technological acceleration decreases long-term risk. As in the policy-free model, acceleration decreases the time spent in any given risky state. Under optimal policy, however, the wealthier future states pulled forward by an acceleration are systematically inclined to be safer, due to the wealth effect. Furthermore, given an increase to the future growth rate, even before actual productive capacity has yet increased, the anticipation of a more valuable future motivates more stringent safety policy in the present.

This analysis might be compared with that of Baranzini and Bourguinion (1995). Baranzini and Bourguinion find conditions under which the growth path that maximizes expected discounted utility also minimizes the probability of existential catastrophe. In our model these objectives never perfectly align, but we explore how technological advances, when regulated with a view to maximizing expected discounted utility, can lower the probability of existential catastrophe.

Sections 2 and 3 explore models in which the *state* of technology at a given time contributes to the hazard rate. Section 4 considers the possibility that risk is "transitional", increasing in the rate of technological *development*.

Absent policy, the effect of acceleration on long-term transition risk is ambiguous. In particular, acceleration has no effect on long-term risk if the "experiment" associated with developing a given technology poses a risk independent of how many experiments happen concurrently. This is the assumption of e.g. Jones's (2016) "Russian roulette" model of risky technological development. If the future contains a sequence of experiments, each of which will pose some risk of catastrophe, then stagnation can lower risk by avoiding advanced experiments altogether; but an acceleration that only pulls forward their date leaves the probability of catastrophe unchanged.

As in Section 3, introducing an optimal policy response facilitates survival due to wealth effects, potentially replacing an ever-increasing hazard rate with a Kuznets curve. Also, though the effect of acceleration on long-term transition risk remains ambiguous given policy, policy can shift the conditions under which acceleration has a given effect on risk. At least in the particular model of transition risk studied in Section 4, the existence of a policy response significantly widens the conditions under which acceleration lowers transition risk.

Section 5 summarizes these analyses and their limitations.

2 State risk without policy

2.1 Model

The hazard rate — The "hazard rate" δ_t is the flow probability at t of anthropogenic existential catastrophe. In this section we posit that it is a positive, continuous function of a state variable A_t:

$$\delta_t = \delta(A_t), \quad \delta(A) > 0 \ \ \forall A.$$

Assume that $A_{(\cdot)}$ is differentiable, with a positive derivative, and that it increases without bound.

We will refer to the state variable as "technology", in acknowledgment of the view that technological developments, broadly construed, are the primary determinants of changes in the hazard rate. In this model, therefore, we proceed through a sequence of overall technology states. A given state may have both risk-inducing features, such as a widespread ability to engineer pathogens, and risk-mitigating features, such as

the ability to easily detect novel diseases, develop vaccines, or implement quarantines. If the "technologies" developed over the period after a state A_t on balance raise the hazard rate, $\delta(A_{t+1}) > \delta(A_t)$. If on balance they lower it, $\delta(A_{t+1}) < \delta(A_t)$.

Survival — The probability that civilization survives to date t is given by

$$S_t \equiv e^{-\int_0^t \delta_s ds} \quad \Longleftrightarrow \quad \dot{S}_t = -\delta_t S_t, \;\; S_0 = 1.$$

The probability that human civilization avoids a catastrophe and, at least in expectation, enjoys a long and flourishing future[3] is

$$S_\infty \equiv \lim_{t\to\infty} S_t = e^{-\int_0^\infty \delta_s ds}. \tag{1}$$

We will refer to $\{\delta_t\}_{t=0}^\infty$ as the *hazard curve*, to the area under the hazard curve $\int_0^\infty \delta_t dt$ as *cumulative risk*, and to S_∞ as the *probability of survival*.

Note that the probability of survival decreases in cumulative risk, and survival is possible ($S_\infty > 0$) iff cumulative risk is finite. Existential catastrophe is not guaranteed only if, roughly, the world is on track eventually to grow ever safer.

2.2 How does acceleration affect risk?

We will now consider how accelerating the overall technology path affects cumulative risk. Importantly, in practice, we may often be able to accelerate or delay particular

[3]In the face of natural existential risk, this will entail succumbing to a natural existential catastrophe instead. From very-long-run historical data on large-scale natural catastrophes, and the typical survival rate of other mammalian species, Snyder-Beattie et al. (2019) estimate that humanity's natural existential hazard rate is below one in 870,000 per year. Throughout this paper we ignore the possibility that technological advances may mitigate natural existential risks. Accounting for this possibility would only strengthen the headline results.

technologies in isolation. Supporting vaccine technologies while discouraging technologies that facilitate engineering pathogens may well lower existential risk from pandemics more than speeding or slowing research across the board. **Such considerations are discussed in Section 3.1.** Here, however, we begin by fixing a path and studying the implications of faster movement along it.

Absent a negative shock severe enough to induce stagnation or recession, technology crosses every value from A_0 to ∞ exactly once. So the area under the hazard curve can be found by integrating with respect to technology instead of time:

$$\int_0^\infty \delta(A_t)dt = \int_{A_0}^\infty \delta(A)\Big(\frac{dA}{dt}\Big)^{-1}dA = \int_{A_0}^\infty \delta(A)\dot{A}_A^{-1}dA, \tag{2}$$

where, somewhat abusing notation, $\dot{A}_A$ denotes the value of $\dot{A}$ when the technology level equals the subscripted A. This change of variables makes it easier to see how various shocks to the growth path affect cumulative risk.

Instantaneous level effects — Consider a shock to the technology level for a short period beginning at t, so that the technology level over this period is approximately $\tilde{A}$ rather than A_t (and the subsequent technology path is unchanged). The sign of the impact of this shock on cumulative risk depends on whether $\delta(\tilde{A})$ is greater or less than $\delta(A_t)$. By the leftmost integral of (2), the impact on cumulative risk *per unit time* of this shock to the technology level at t equals

$$\delta(\tilde{A}) - \delta(A_t).$$

Instantaneous accelerations — Consider the impact on cumulative risk per unit time of an instantaneous shock to technology *growth* at t, so that the technology growth

rate at t is $\dot{\tilde{A}}$ rather than $\dot{A}_t$, and the subsequent technology growth rate at each technology level is unchanged. By the rightmost integral of (2), the impact of this shock on cumulative risk *per unit of increase to the technology level* during the acceleration is $\delta(A_t)(\dot{\tilde{A}}^{-1} - \dot{A}_t^{-1})$. Multiplying this by the new rate of technology growth per unit time, the impact on cumulative risk per unit time equals

$$-\delta(A_t)\big(\dot{\tilde{A}}/\dot{A}_t - 1\big).$$

Accelerations — Choose technology level $\overline{A} > A_t$. Since the baseline technology path increases continuously and without bound, $\overline{A} = A_T$ for some $T > t$.

Consider the effect of increasing the technology level at t from A_t to $\overline{A}$ and subsequently maintaining the technology path $\overline{A}_s = A_{s+(T-t)}$ $(s \geq t)$. This shock to the technology path amounts to a "leap forward in time", cutting a slice cut out of the hazard curve. Cumulative risk falls by

$$\int_{A_t}^{\overline{A}} \delta(A)\dot{A}_A^{-1}\, dA.$$

More generally, define a *temporary acceleration* as an increase to $\dot{A}$ at some range of technology levels: say, from A_t to A_T. Because the exponent on $\dot{A}$ in the integral is negative, the acceleration lowers the risk endured at the given range of technology levels. A discontinuous jump in the technology level amounts to raising $\dot{A}_A$ to ∞, and thus lowering $\dot{A}_A^{-1}$ to 0, from $A = A_t$ to A_T.

A jump in the technology level from A_t to A_T temporarily increases the hazard rate if $\delta(A_T) > \delta(A_t)$. Likewise, an acceleration to technology growth accelerates an increase to the hazard rate if $\delta(\cdot)$ is increasing around A_t. It may therefore appear

to contemporaries that a given permanent level effect decreases the probability of survival. Here, that would be incorrect. If (2) is infinite, the probability of survival is zero with or without the permanent level effect.[4] If (2) is finite, the permanent level effect decreases cumulative risk and increases the probability of survival.

Define a *permanent acceleration* to be a permanent increase to $\dot{A}$ from some time t—or, equivalently, some technology level A_t—onward. By the rightmost integral of (2), a permanent acceleration, like a temporary acceleration, must lower cumulative risk if cumulative risk is finite on the baseline technology path.

Unlike temporary accelerations, however, permanent accelerations can render survival possible when it would otherwise be impossible. Shrinking a heavy-tailed curve with an infinite integral can yield a thin-tailed curve with a finite integral.

To state this lesson in reverse, consider stagnation: a permanent negative acceleration, or "deceleration", setting $A_s = A_t$ for all $s \geq t$. The hazard rate is then permanently positive, and survival is impossible, even if it might have been possible with technology growth. More concretely, consider the implications of a large negative shock today returning the world to the state it inhabited in 1924. Perhaps the hazard rate was much lower in 1924 than today, but this reset would largely doom us to relive the nuclear standoffs, emissions-intensive industrializations, and biotechnological hazards of the past. With enough replays of the past century, a catastrophe would presumably be inevitable.

[4] $\int_{A_t}^{A_T} \delta(A_t)dt$ is finite by the continuity of δ in A and of A in t.

3 State risk with policy

3.1 Motivation

The previous section shows that, under weak conditions, acceleration along a "technology path" either lowers or does not affect cumulative risk. Any optimism one might draw from this observation about the risk implications of accelerating technological development in practice faces two limitations.

First, cumulative risk falls only if we are already on track to eventually grow ever safer. Because the technology path was assumed to be fully exogenous, the model of the previous section offers no reason to believe we are. If technological progress, broadly construed, has historically increased the hazard rate, the message of Section 2 is that those who wish to reduce existential risk should accelerate such progress on the blind hope that the associated risk trend eventually reverses.

Second, the result only applies to a perfectly balanced acceleration in the path of every determinant of the hazard rate. It says nothing about the implications of pulling forward some technologies in time without equally pulling forward all others.

To illustrate the second limitation, suppose the hazard rate is a function of two variables, A and x, and suppose the path of x is fixed:

$$\delta_t = A_t x_t, \quad x_t = (1+t)^{-2}.$$

Here, later technology states A_t are always riskier on balance. x_t may represent an index of policies and tools which promote safety ever more effectively with time and are independent the state of (acceleratable) technology. Consider an acceleration in the technology path from $A_t = (1+t)^k$ to $A_t = (1+t)^{\tilde{k}}$, where $k < 1 < \tilde{k}$. This

acceleration increases cumulative risk from

$$\int_0^\infty (1+t)^{k-2}dt \quad \text{to} \quad \int_0^\infty (1+t)^{\tilde{k}-2}dt.$$

The former is finite, because $k-2<-1$. The latter is infinite, because $\tilde{k}-2>-1$. In this case, accelerating A lowers the probability of survival to zero.

At the other extreme, however, suppose that the index x_t of risk-relevant features of the world not included in a given technological acceleration is set by policy, in light of the path of A, to optimize a tradeoff between safety and consumption. As long as more advanced technology levels make greater *consumption* feasible, we will see that optimal policy—with respect to any discount rate—generally *strengthens* the conclusion that technological acceleration lowers cumulative risk.

As in the tech-only model of Section 2, survival can only be achieved by pulling forward a future that asymptotically approaches perfect safety. Whereas the earlier model is agnostic about whether more advanced technology will in fact carry a lower hazard rate, however, an optimal policy response introduces a tendency for faster technological development to carry lower risk in the long run. This is because technology increases consumption, which both decreases the utility cost of a marginal consumption sacrifice and increases the value of life. Furthermore, the prospect of a *future* acceleration now lowers the *present* hazard rate, because when the value of the future is greater, it is worth sacrificing more today to prevent its ruin.

These dynamics are illustrated in a simple model of technology and optimal policy in the rest of this section. **Generalized results, not relying on functional form assumptions, are given in Appendix A.3.**

3.2 The economic environment

Technology — The maximum feasible consumption level at t equals the technology level A_t. Actual consumption, C_t, is A_t multiplied by a policy choice $x_t \in [0, 1]$:

$$C_t = A_t x_t. \tag{3}$$

The tradeoff at the heart of this section is that a technologically advanced civilization can risk self-destruction, but that this risk can be lowered at some cost to consumption, as represented here by a choice of x below 1. (We denote the choice variable x to remind the reader that higher choices of x come with higher existential risk.) Choices of x below 1 may constitute bans on the adoption of risky production processes and/or resource-allocations to safety-increasing services like pandemic detection.

The technology frontier A grows at a constant rate g:

$$\dot{A}_t = A_t g, \quad g > 0, A_0 > 1. \tag{4}$$

The hazard rate — The hazard rate δ_t is now a function of the technology level A_t and the policy choice $x_t \in [0, 1]$, and is increasing in x_t. In this illustrative model, the elasticities of the hazard rate in A and in x are constant:

$$\delta(A_t, x_t) = \bar{\delta} A_t^{\alpha} x_t^{\beta}, \quad \bar{\delta} > 0, \ \beta > \alpha > 0, \ \beta > 1. \tag{5}$$

We impose the three inequalities of $\beta > \alpha > 0$, $\beta > 1$ to satisfy three desiderata.[5]

[5]Hazard function (5) is closely analogous to the environmental damage function of Stokey (1998). While Stokey focuses on the implications of the damage function for the chosen path of x (or "z" in her notation), we will study how accelerations to the path of A affect the probability of a binary event: the occurrence of an anthropogenic existential catastrophe at any time.

The first is that, fixing $x_t > 0$, δ_t increase in A_t. This imposes $\alpha > 0$. The assumption that δ_t increase in A_t is necessary if we are to concede that technological development has historically increased the hazard rate, and that this trend would continue absent a change in policy.[6] It is thus necessary to illustrate that optimal policy can render survival possible when it might otherwise have been impossible.

Second, the elasticity of δ_t with respect to x_t is assumed to exceed its elasticity with respect to A_t; i.e. $\beta > \alpha$. This is equivalent to the condition that, when technology advances, it is feasible to lower the hazard rate by retaining the former consumption level, allocating all marginal productive capacity to safety measures. This may be seen by substituting $x_t = C_t/A_t$ (from (3)) into the hazard function (5), yielding

$$\delta_t = \bar{\delta} A_t^{\alpha-\beta} C_t^{\beta}.$$

Fixing C, the hazard rate falls over time iff $\beta > \alpha$. If it is indefinitely infeasible to lower the hazard rate while fixing consumption, as it is in this model if $\beta \leq \alpha$, then an existential catastrophe is unavoidable unless consumption falls to zero. This degrowth would amount to the destruction of civilization by other means. If $\beta \leq \alpha$, therefore, speeding or slowing growth can have no impact on the probability of an existential catastrophe broadly construed.

Third, fixing $A_t > 0$, $\beta > 1$ renders δ_t strictly convex in x_t, so that there are

[6]The proportion $1 - x$ of potential consumption sacrificed for the sake of existential safety has only increased as technology has advanced. Ord (2020, p. 313) estimates that, as of 2020, approximately \$100M/year was spent specifically on reducing existential risk. This is likely a great underestimate of existential safety expenditures in the sense relevant here, for two reasons. First, explicit expenditures do not include foregone consumption due to regulatory barriers. Second, many catastrophic risk reduction efforts are motivated both by the desire to reduce existential risks and by the desire to reduce smaller-scale damages. By contrast, Moynihan (2020) argues that the very concept of an anthropogenic existential catastrophe essentially did not exist 300 years ago; it appears there were then no efforts taken to prevent one.

diminishing returns to safety efforts. We consider this assumption reasonable both from first principles and from Shulman and Thornley's (2024) estimates of the cost-effectiveness of existential risk mitigation efforts (Appendix A.1).

Preferences — A planner maximizes expected discounted flow utility, which is a CRRA function of consumption:

$$\int_0^\infty e^{-\rho t} S_t \, u(C_t) \, dt; \qquad u(C_t) = \begin{cases} \frac{C_t^{1-\gamma}-1}{1-\gamma}, & \gamma > 0, \neq 1; \\ \log(C_t), & \gamma = 1. \end{cases} \tag{6}$$

The discount rate $\rho > 0$ is the sum of some rate of pure time preference and/or some rate of natural and unavoidable existential risk.[7] When $\gamma < 1$, we will impose

$$\rho > \underline{\rho} \equiv \frac{(\beta - \alpha)(1-\gamma)}{\beta} g \tag{7}$$

to ensure the existence of an optimal policy path.

The utility of death is implicitly normalized to 0 and the death-equivalent consumption level to 1. Equivalently, we are normalizing to 1 the technology level at which, when consumption is maximized, flow utility equals 0.

Note that when $\gamma > 1$, flow utility is upper-bounded by $\frac{1}{\gamma-1}$. Accelerating consumption growth, from a baseline of positive consumption growth, therefore yields a stream of utility increases that eventually shrinks over time. Concern for the future then clearly casts doubt on the value of speeding technological development: the consumption benefits of doing so primarily accrue in the short run, whereas the costs

[7] One valid interpretation of these preferences would be that the population is fixed and (6) is the expected utility of a representative household. Another would be that population grows exponentially at rate $n < \rho$, that the rate of pure time preference and exogenous risk is in fact $\rho + n$, and that the planner uses the total utilitarian social welfare function.

of an existential catastrophe are everlasting. By contrast, when $\gamma \leq 1$, flow utility can grow without bound, so accelerations to consumption growth and reductions in existential risk can have comparable long-term benefits.

3.3 The existential risk Kuznets curve

Optimality — Summarizing Section 3.2, the planner chooses a policy path $\{x_t\}_{t=0}^{\infty}$ to maximize expected utility (6) subject to

$$\begin{aligned} &A_0 > 1, \;\; \dot{A}_t = gA_t \;\; (g > 0), \\ &C_t = A_t x_t, \\ &S_0 = 1, \;\; \dot{S}_t = -\delta_t S_t, \\ &\delta_t = \bar{\delta} A_t^{\alpha} x_t^{\beta} \;\; (\bar{\delta} > 0, \; \beta > \alpha > 0, \; \beta > 1). \end{aligned} \tag{8}$$

This section finds the path of the hazard rate in the planner's solution. The next section explores what this implies for the impact of acceleration on cumulative risk.

The planner faces one choice variable, x_t, and one state variable, S_t. Her expected flow utility at t is $S_t u(C_t)$. Her problem is represented by current-value Lagrangian

$$\begin{aligned} \mathcal{L}_t &= S_t u(C_t) + v_t \dot{S}_t + \mu_t (1 - x_t) \\ &= S_t \frac{(A_t x_t)^{1-\gamma} - 1}{1-\gamma} - v_t \, \bar{\delta} A_t^{\alpha} x_t^{\beta} S_t + \mu_t (1 - x_t). \end{aligned} \tag{9}$$

μ_t is the Lagrange multiplier on x, positive iff the $x_t \leq 1$ constraint binds.

$$v_t = \int_t^{\infty} e^{-\rho(s-t)} \frac{S_s}{S_t} u(C_s) ds \tag{10}$$

is the costate variable on survival: the expected value of civilization as of t.[8]

On an optimal path, the first-order condition on (9) with respect to the choice variable x_t is satisfied. Differentiating (9) with respect to x_t, we have

$$S_t A_t^{1-\gamma} x_t^{-\gamma} - \bar{\delta} A_t^{\alpha} \beta x_t^{\beta-1} v_t S_t \geq 0, \tag{11}$$

with inequality iff the left-hand side is positive at $x_t = 1$, in which case $x_t = 1$ is optimal.[9] So as long as (11) is nonnegative at $x_t = 1$, the optimal $x_t \in [0, 1]$ equals 1. Any consumption sacrifices would carry greater flow costs than expected benefits. When (11) is negative at $x_t = 1$, the optimal choice of x_t is interior. It sets (11) equal to zero, maintaining that on the margin, the loss of flow utility from lowering consumption equals the expected benefit via risk reduction.[10]

In fact there is a unique[11] optimal path, characterized by (11), a first-order condition corresponding to S_t, and identity (10) (see Appendix B.1).

Initial risk increases — (11) is nonnegative at $x_t = 1$ iff

$$A_t^{-(\alpha+\gamma-1)} \geq \bar{\delta} \beta v_t. \tag{12}$$

The continuation value of civilization at t given survival to t, v_t, strictly rises over time. This is because, given the best paths $\{C_s\}_{s \geq t}$ and $\{\delta_s\}_{s \geq t}$ achievable at a given initial technology level A_t, a higher initial technology level allows for a path with an equal hazard rate but more consumption at each future period, given $\beta > \alpha$. A higher initial technology level makes a preferred future feasible.

[8] The costate variable on survival equals (10) because the value of saving the world equals the expected value of the world. The equation is derived formally in Appendix B.1.

[9] The second derivative with respect to x_t is negative because $\beta > 1$.

[10] We can ignore the possibility that optimal x_t equals 0 because this yields infinite flow disutility.

[11] Given semi-continuity. If path x is optimal, measure-zero deviations from x are also optimal.

Suppose inequality (12) is satisfied strictly at $t = 0$. Then early in time, when A_t is low, the optimal policy choice is $x = 1$, and the hazard rate rises at rate[12]

$$g_{\delta t} = \alpha g.$$

Eventual risk declines and survival — As the left-hand side of (12) falls exponentially with A_t and the right-hand side rises, there is a unique time t^* at which (12) holds with equality. After t^*, the optimal choice of x_t is interior and sets (11) equal to zero.

Setting (11) equal to zero and rearranging, we have the optimal choice of x_t after t^*, and thus the optimal choice of x_t in general:

$$x_t = \begin{cases} 1, & t \leq t^*; \\ \left(\bar{\delta}\beta A_t^{\alpha+\gamma-1} v_t\right)^{-\frac{1}{\beta+\gamma-1}}, & t > t^*. \end{cases} \tag{13}$$

Taking growth rates, we find the growth rate of the policy variable after t^*:

$$g_{xt} = -\frac{\alpha+\gamma-1}{\beta+\gamma-1} g - \frac{1}{\beta+\gamma-1} g_{vt}. \tag{14}$$

The hazard rate in turn grows as

$$g_{\delta t} = \alpha g + \beta g_{xt} = -\frac{(\beta-\alpha)(\gamma-1)}{\beta+\gamma-1} g - \frac{\beta}{\beta+\gamma-1} g_{vt}. \tag{15}$$

[12]Given a time-dependent variable y, $g_{yt} \equiv \dot{y}_t / y_t$ denotes its proportional growth rate at t.

Finally, on an optimal path, v grows asymptotically at a constant rate, with

$$\lim_{t\to\infty} g_{vt} = \begin{cases} \frac{(\beta-\alpha)(1-\gamma)}{\beta} g, & \gamma < 1; \\ 0, & \gamma \geq 1. \end{cases} \tag{16}$$

This is proved in Appendix B.2, but an intuition is as follows. When $\gamma > 1$, growth in v must fall to zero because v_t is upper-bounded by

$$\bar{v} \equiv \frac{1}{\rho(\gamma - 1)}. \tag{17}$$

When $\gamma < 1$, flow utility grows approximately like $C_t^{1-\gamma}$ when C_t is large. Observe from (10) that v_t grows roughly with flow utility. Substituting $g_v = (1-\gamma)g_C = (1-\gamma)(g + g_x)$ into (15) then yields the $\gamma < 1$ case of (16).

Substituting (16) into (14) and (15) gives the asymptotic growth rates g_x and g_δ. Also, since $C_t = A_t x_t$, we can easily find the asymptotic value of g_C, which is always positive. Though x falls to 0, A grows more quickly than x declines. Indeed, as elaborated in Appendix A.3.3, consumption growth is key to the growth in sacrifices for safety. With decreasing marginal utility to *consumption* and decreasing marginal returns to *sacrifices for safety*, potential consumption is split between the former and latter so that the marginal value of each stays equal.

Proposition 1. ***The existential risk Kuznets curve***

On the path defined by (6)–(8), there is a time $t^* \geq 0$ such that for $t < t^*$,

$$x_t = 1, \quad g_{Ct} = g > 0, \quad g_{\delta t} = \alpha g > 0;$$

and for $t > t^*$ x_t is interior, such that if $\gamma > 1$,

$$\lim_{t\to\infty} g_{xt} = -\frac{\alpha+\gamma-1}{\beta+\gamma-1}g < 0, \tag{18}$$

$$\lim_{t\to\infty} g_{Ct} = \frac{\beta-\alpha}{\beta+\gamma-1}g > 0, \tag{19}$$

$$\lim_{t\to\infty} g_{\delta t} = -\frac{(\beta-\alpha)(\gamma-1)}{\beta+\gamma-1}g < 0; \tag{20}$$

and if $\gamma \leq 1$,

$$\lim_{t\to\infty} g_{xt} = -\frac{\alpha}{\beta}g < 0, \tag{21}$$

$$\lim_{t\to\infty} g_{Ct} = \frac{\beta-\alpha}{\beta}g > 0, \tag{22}$$

$$\lim_{t\to\infty} \delta_t t = \frac{\rho}{(\beta-\alpha)g} > 0, \;\; \gamma = 1; \tag{23}$$

$$\delta^* \equiv \lim_{t\to\infty} \delta_t = \frac{(\rho-\underline{\rho})(1-\gamma)}{\beta+\gamma-1} > 0, \;\; \gamma < 1. \tag{24}$$

Proof. See Appendix B.2. □

Corollary 1.1. ***Survival***

$S_\infty > 0$ iff $\gamma > 1$.

Proof. The "if" follows from (20) and the definition of S_∞. δ_t ultimately falls exponentially, so $\int_0^\infty \delta_t dt < \infty$, so $S_\infty \equiv e^{-\int_0^\infty \delta_t dt} > 0$. The "only if" likewise follows from (23)–(24): when the hazard rate is asymptotically constant, or falls proportionally to $1/t$ (or more slowly), the integral of the hazard curve diverges. □

Intuition for the $\gamma = 1$ threshold — The importance of γ stems from the fact that, for the policy path to be optimal, it must maintain

a) the flow utility to proportionally increasing consumption, $C_t \cdot C_t^{-\gamma}$

=

b) the damage done via proportionally raising the hazard rate,

which equals the hazard rate × the value of civilization.

When the value of civilization also grows like $C_t^{1-\gamma}$, as it does when $\gamma < 1$, the hazard rate must be constant for (a) and (b) to grow at the same rate. When $\gamma > 1$, the value of civilization is asymptotically constant, so the hazard rate falls like $C_t^{1-\gamma}$. When $\gamma = 1$, given that consumption grows exponentially, $\log(C_t)$ and thus v_t grow linearly. The hazard rate then falls proportionally to $1/t$.

This result recalls the "Russian roulette" model of Jones (2016). There, optimal policy renders catastrophe avoidable iff $\gamma \geq 1$. Since risk in that model is posed by the development (not existence) of technologies, it more similar to the "transition risk" model of Section 4, and is discussed further there.

Simulation — The paths of policy and risk are simulated below, for the parameter values listed in Table 1.

ρ	γ	g	A_0	α	β	δ
0.02	1.5	0.02	2	1	2	0.00012

Table 1: Simulation parameters for Figure 1

The values of ρ, γ, and g have been chosen as central estimates from the macroeconomics literature.[13] $A_0 = 2$ is chosen so that the value of a statistical life-year at $t = 75$ is 4x consumption per capita, roughly matching Klenow et al. (2024).[14] That

[13]We focus on the $\gamma > 1$ case both because it generates the important results and to match evidence from Hall (1988), Lucas (1994), Chetty (2006), and others.

[14]They estimate that this ratio was roughly 5 in the United States in 2019. The figure must be adjusted upward for economic growth since 2019, but downward insofar as we are considering optimal policy across all countries advanced enough to be deploying existentially hazardous technology.

is, the first year of the simulation might be taken to denote 1949, when a nuclear war between superpowers first became possible, in which case the 75th year denotes the present. $\bar{\delta}$, α, and β are chosen so that the hazard rate today is approximately 0.1%, matching Stern's (2007) oft-cited figure; and, for clarity in illustration, so that the hazard rate begins to fall at approximately $t = 100$, and so that the growth rate and then the decay rate of the hazard rate are non-negligible.

S_∞ under these parameters, from $t = 75$ onward, is approximately 65%.

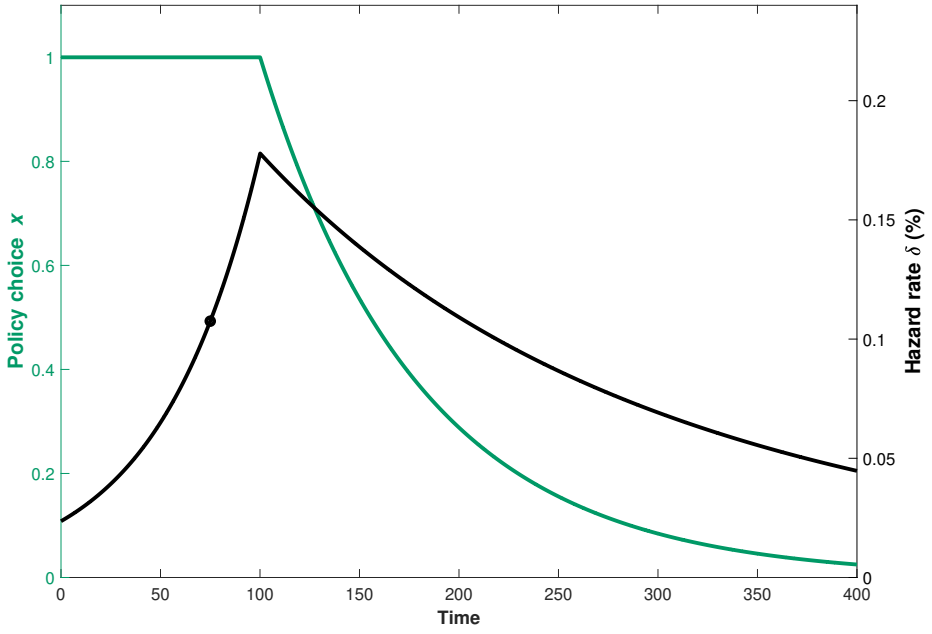


Figure 1: Evolution of the policy choice and the hazard rate along the optimal path

Calculations and code for replicating the simulation and corresponding probability of survival may be found in Appendix C.

As Figure 1 illustrates, one potentially unappealing feature of this simple model is that it implies that, on the optimal path, the hazard rate only rises while no sacrifices whatsoever are made for existential safety. In this it resembles Stokey's

(1998) "environmental Kuznets curve", whose damages also rise exponentially with growth and then fall sharply once it becomes optimal to take action.

As in Stokey (1998), this dynamic is driven by the lack of a lower Inada condition on $1 - x$. If marginal "safety expenditures" lower the hazard rate infinitely per unit spent at $x = 1$, then as long as $v_t > 0$ it is optimal to set $x_t < 1$, even if at first the hazard rate is allowed to rise. Rising δ can thus be found alongside falling x by tweaking the hazard function around $x = 1$. Such tweaks do not affect the long-run behavior of policy or risk as given by (18)–(20), which are set by the shape of the hazard function around $x = 0$. This is discussed further in Appendix A.4.1.

3.4 Acceleration and state risk

As in the tech-only model of Section 2, the impact on cumulative risk of a temporary shock is ambiguous, but the impact of an acceleration—e.g. a permanent level or growth effect—is always weakly to lower cumulative risk.

Preliminaries — Let $A_{(\cdot)}$ denote the baseline technology path, given by (4). Let $A^* \equiv A_{t^*}$, where t^* is defined as in Proposition 1.

The area under the hazard curve can again be defined by integrating with respect to A instead of t. We will let X denote cumulative risk given that the technology path is $A_{(\cdot)}$ and the policy path x is optimal given $A_{(\cdot)}$:

$$X \equiv \int_0^\infty \bar{\delta} A_t^\alpha x_t^\beta dt = \int_{A_0}^\infty \bar{\delta}\, A^\alpha\, x_A^\beta \left(\frac{dA}{dt}\right)^{-1} dA = \int_{A_0}^\infty \bar{\delta}\, A^\alpha\, x_A^\beta\, \dot{A}_A^{-1}\, dA, \tag{25}$$

where we will again abuse notation somewhat by letting x_A and $\dot{A}_A$ denote, respectively, the optimal value of x (given technology path $A_{(\cdot)}$) and the value of $\dot{A}$ when the technology level equals the subscripted A.

We will define v_A and δ_A likewise. Note that $\delta_A \equiv \bar{\delta} A^{\alpha} x_A^{\beta}$, without dividing this expression by $\dot{A}_A$. That is, δ_A is still a hazard rate: it denotes the probability of catastrophe per unit time at technology level A, not the probability of catastrophe per unit of technological development.

A technology path $\tilde{A}_{(\cdot)}$ that is continuously differentiable almost everywhere and whose right derivative $\dot{\tilde{A}}_{(\cdot)}$ is defined and right-continuous everywhere is an *acceleration* from $\underline{A} \in [A_0, \infty)$ to $\overline{A} \in (A_t, \infty]$ if $\tilde{A}_0 = A_0$ and

$$\dot{\tilde{A}}_A = \dot{A}_A, \;\; A \notin (\underline{A}, \overline{A}); \qquad \dot{\tilde{A}}_A > \dot{A}_A, \;\; A \in (\underline{A}, \overline{A}).$$

The acceleration is *permanent* if $\overline{A} = \infty$ and *temporary* otherwise.

Let $\tilde{A}_{(\cdot)}$ be an acceleration from $\underline{A}$. Define $\tilde{v}_A$ such that at $A < \underline{A}$, $\tilde{v}_A = v_A$, and at $A \geq \underline{A}$, $\tilde{v}_A$ is the costate variable on survival at A given that the subsequent technology path is $\tilde{A}_{(\cdot)}$. Then $\tilde{x}_A$ is defined to equal (13) with $A, \tilde{v}_A$ in place of A_t, v_t; $\tilde{\delta}_A \equiv \delta(A, \tilde{x}_A)$; and $\tilde{X} \equiv \int_{A_0}^{\infty} \tilde{\delta}_A \dot{\tilde{A}}_A^{-1} dA$.

Given a baseline technology level $\underline{A}$ and a technology growth rate $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$, denote by $\tilde{A}_{(\cdot)}[\epsilon]$ the acceleration from $\underline{A}$ to $\underline{A} + \epsilon$ with

$$\dot{\tilde{A}}_A = \dot{\tilde{A}}, \quad A \in (\underline{A}, \underline{A} + \epsilon).$$

Then the effect on cumulative risk, per unit of technological development, of *instantaneously accelerating to* $\dot{\tilde{A}}$ *at* $\underline{A}$ is defined to be

$$\Delta_{\underline{A}, \dot{\tilde{A}}} \equiv \lim_{\epsilon \to 0} \left(\tilde{X}[\epsilon] - X \right) / \epsilon,$$

where $\tilde{X}[\epsilon]$ is cumulative risk $\tilde{X}$, as defined above, given acceleration $\tilde{A}_{(\cdot)}[\epsilon]$.[15]

Instantaneous level effects — The effect per unit time of a positive shock to the technology level A_t, letting policy adjust instantaneously, depends on whether the shock occurs before or after the regime-change time t^*. At $t < t^*$, temporarily multiplying the technology level by $m > 1$ has no impact on the optimal choice of x.[16] The hazard rate thus rises. The future hazard rate is unaffected, so cumulative risk increases by

$$\delta_t(m^\alpha - 1) > 0$$

per unit of time that the technology level is raised.

At $t \geq t^*$, temporarily multiplying the technology level by $m > 1$ multiplies the policy variable by $m^{-\frac{\alpha+\gamma-1}{\beta+\gamma-1}}$, by (13). In combination, the positive shock to technology and the negative impact on the policy variable multiply the hazard rate by $m^{\alpha-\beta\frac{\alpha+\gamma-1}{\beta+\gamma-1}} = m^{-\frac{(\beta-\alpha)(\gamma-1)}{\beta+\gamma-1}} < 1$. This resulting change in cumulative risk is

$$\delta_t\big(m^{-\frac{(\beta-\alpha)(\gamma-1)}{\beta+\gamma-1}} - 1\big) < 0$$

per unit of time that the technology level is raised.

Instantaneous accelerations — Multiplying the technology growth rate at t by $m > 1$ lowers cumulative risk (per unit of time that the shock lasts) regardless of t. It does so only because the shock decreases the time spent at technology levels around A_t. The shock has no impact on the policy associated with any technology level.

---

[15] The effect on an instantaneous acceleration on cumulative risk *per unit time* is $\Delta_{\underline{A},\dot{A}}\dot{\tilde{A}}$, since during the acceleration, $\dot{\tilde{A}}$ units of technology are developed per unit time. This is of the same sign.

[16] Unless m is large enough to reverse inequality (12), a case we will ignore for simplicity.

As in the tech-only model, therefore, the impact of this shock on cumulative risk *per unit of increase to the technology level* during the acceleration is

$$\delta_t((m\dot{A}_t)^{-1} - \dot{A}_t^{-1}) < 0.$$

So the impact on cumulative risk *per unit of time* that the acceleration lasts is the above multiplied by the new technology growth rate $m\dot{A}_t$:

$$\delta_t(1 - m) < 0.$$

Accelerations — Consider a "sharp temporary acceleration", in which technology jumps at t from A_t to $\overline{A} > A_t$ and exponential technology growth is then maintained. Since in this model optimal policy is history-independent, this technology shock amounts to a leap forward in time. The resulting change in cumulative risk is

$$-\int_{A_t}^{\overline{A}} \delta_A \dot{A}_A^{-1}\, dA.$$

Such a jump is the limiting case of an acceleration from $\underline{A}$ to $\overline{A}$, which can lower the risk endured at the given range of technology levels for two reasons. First, as in the tech-only model, increasing the technology growth rate at A always lowers cumulative risk directly because the exponent on $\dot{A}_A$ in integral (25) is negative: $\tilde{\dot{A}}_A^{-1} < \dot{A}_A^{-1}$. Second, going beyond the tech-only model, given $A \in [A_t, \overline{A})$, the value of the future at A is higher given faster future technology growth: $\tilde{v}_A > v_A$. By (13), this motivates weakly more stringent policy $\tilde{x}_A \leq x_A$ and thus a weakly lower hazard rate $\tilde{\delta}_A \leq \delta_A$.

Via the first channel alone, the change in cumulative risk achieved by an accel-

eration is the integral, across technology levels, of the risk reductions achieved by instantaneous accelerations at each technology level:

$$\int_{A_t}^{\overline{A}} \delta_A(\dot{\tilde{A}}_A^{-1} - \dot{A}_A^{-1})dA < 0.$$

Given a policy impact, the cumulative risk reduction achieved is greater.

To summarize:

Proposition 2. ***Acceleration and state risk***

If $\gamma > 1$, so that X is finite, an instantaneous acceleration at $\underline{A}$ to $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$ decreases cumulative risk per unit of technological development during which it endures:

a) $\Delta_{\underline{A},\dot{\tilde{A}}} = \delta_{\underline{A}}(\dot{\tilde{A}}^{-1} - \dot{A}_{\underline{A}}^{-1}) < 0$

and an acceleration $\tilde{A}_{(.)}$ from $\underline{A}$ to $\overline{A}$ decreases cumulative risk by weakly more than the corresponding integral of instantaneous accelerations:

b) $\tilde{X} \leq X + \int_{\underline{A}}^{\overline{A}} \Delta_{A,\dot{\tilde{A}}_A} dA < X$, *with equality strict only if* $\overline{A} \leq A^*$.

The results follow from the integral defining cumulative risk (25) and the definition of instantaneous acceleration. The impacts of shocks to growth on survival are explored in the generalized model of Appendix A.3, and the generalized results are given and proved in detail there in Proposition 7.

Simulation — The effects of a "sharp temporary acceleration" are illustrated in Figure 2. The parameter values used to illustrate the baseline path are the same as those used to simulate Figure 1. The acceleration takes place "today", at $t = 75$, and multiplies A by $e^{0.2} \approx 1.22$, so that at $g = 0.02$, it amounts to a 10-year leap forward.

Recall from the simulation of the previous section that the probability of survival (from $t = 75$ onward) on the baseline path is approximately 65%. The proportional increase in the probability of survival can be found analytically. Cumulative risk X declines by precisely the area under the baseline hazard curve from $t = 75$ to 85; and since $\delta_{75} = 0.1\%$, $g = 0.02$, and $\alpha = 1$, this difference equals

$$\Delta X = -0.001 \int_0^{10} e^{0.02t} dt = -0.05(e^{0.2} - 1).$$

$S_\infty = e^{-X}$ is then multiplied by $e^{-\Delta X} \approx 1.011$, so that in absolute terms S_∞ rises by approximately $0.65 \cdot 0.011 \approx 0.7\%$.

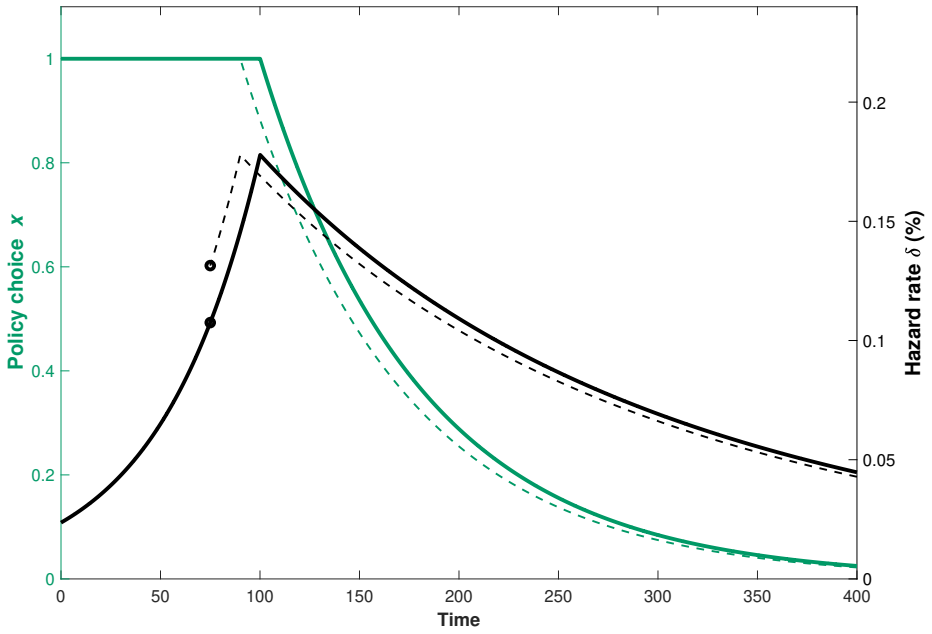


Figure 2: Acceleration shrinks cumulative risk

Calculations and code for replicating the simulation may be found in Appendix C.

3.5 Discussion

Slow growth makes catastrophe inevitable — As noted in Section 2.2, "deceleration" can render survival impossible. For example, stagnation must do so.

Here, the technology conditions necessary for survival can be stated more precisely. Suppose $\gamma > 1$ and consider a permanent deceleration after which technology grows power-functionally, so that $\tilde{A}_t = t^k$ for some $k > 0$. The exponential growth rate of $\tilde{A}$, denoted $\tilde{g}$, is then time-varying, with $\tilde{g}_t = k/t$. By (15) and since $\tilde{g}_v \to 0$, δ_t then falls to 0 like $t^{-\frac{(\alpha-\beta)(\gamma-1)}{\beta+\gamma-1}k}$. Since cumulative risk is finite for $\delta_t \propto t^{-\kappa}$ iff $\kappa > 1$, the probability of survival is positive iff

$$k > \frac{\beta + \gamma - 1}{(\alpha - \beta)(\gamma - 1)}. \tag{26}$$

Growth vs. patience — Faster growth increases the willingness to pay for safety. Those concerned about the safety of the long-term future often pursue the same result via ethical arguments for a low rate of time preference. Consider e.g. the Stern–Nordhaus debate (and the long debate since) over the discount rate to use in climate policy, or the arguments for concern for the future made by philosophers such as Parfit (1984), Cowen and Parfit (1992), Ord (2020), and MacAskill (2022).

Mere level effects—permanent proportional increases to A—here impact policy and survival probability similarly to proportional decreases to ρ. In brief, this is because proportionally decreasing ρ raises v by a similar proportion (c.f. (17)), and by (13), proportional increases to v and to A have similar effects on policy.

Acceleration can lower life expectancy — If $\gamma < 1$, so that civilization's "life expectancy" is finite, accelerations can decrease it. To see this, recall that stagnation

at low A yields a permanent hazard rate of $\bar{\delta}A^{\alpha}$. This may be arbitrarily low, so the expected duration until catastrophe $1/(\bar{\delta}A^{\alpha})$ may be arbitrarily high. When $\gamma < 1$, an acceleration can quickly yield a hazard rate that permanently approximates δ^* (24). Civilizational life expectancy can thus fall to approximately $1/\delta^*$.

4 Transition risk

4.1 Motivation

A hazard function of the form $\delta(A_t, x_t)$ captures what we have called "state risk": δ depends on the *level* of technology. On this framing, it is perhaps unsurprising that escaping risky states more quickly lowers cumulative risk.

But risk may instead be "transitional": posed by *technological development*. This is the intuition captured by Jones's (2016) "Russian roulette" model of technological development and (2024) model of AI risk, and by Bostrom's (2019) analogy to drawing potentially destructive balls from an urn. Perhaps stagnation at a given level of technology is essentially safe, and risk is posed by discovering and deploying new technologies with unknown consequences. If so, given a positive-growth baseline, does accelerating technological development further increase cumulative risk?

4.2 A transition-risk-based hazard function

To explore this possibility, suppose δ increases in $\dot{A}_t$ instead of, or as well as, in A_t. We will again restrict our consideration to a constant elasticity hazard function:

$$\delta_t = \bar{\delta}A_t^{\alpha}\dot{A}_t^{\zeta}x_t^{\beta}, \quad \bar{\delta} > 0,\ \zeta \geq 0,\ \beta > 1. \tag{27}$$

Hazard function (5) is the special case of (27) with $\zeta = 0$ (and $\beta > \alpha > 0$). If $\zeta > 0$, however, the model is most naturally interpreted as one in which risk is posed by the introduction of new technologies—"draws from Bostrom's urn"—which each increase A by one unit. Fixing policy, introducing multiple technologies can pose more, less, or equal risk if done concurrently than if done in sequence, depending on the sign of $\zeta - 1$. Introducing more advanced technologies can pose more, less, or equal risk than introducing less advanced technologies, depending on the sign of α.

Alternatively, to interpret one "new technology" as a *proportional* increase to A, simply rewrite the hazard function as

$$\delta_t = \bar{\delta} A_t^{\alpha+\zeta} \left(\frac{\dot{A}_t}{A_t}\right)^{\zeta} x_t^{\beta}.$$

Here $\alpha + \zeta > 0$ is the condition under which developing more advanced technologies poses more risk than developing less advanced technologies. If $\dot{A}/A$ has long been roughly constant, the view that the hazard rate has risen must be attributed to the increasing danger of each "technological development" in this sense.

Finally, consider the case of $\alpha = -1$, $\zeta = 1$, so that

$$\delta_t = \bar{\delta} \frac{\dot{A}_t}{A_t} x_t^{\beta}.$$

Here, fixing x, each proportional increase to A induces the same hazard, independently of how quickly the increase occurs. In the absence of policy—with $x = 1$ (or any other constant) permanently—this model is essentially equivalent to the "Russian roulette" model of Jones (2016)[17] and the AI risk model of Jones (2024).

[17] Our $\bar{\delta}$ is the variable there denoted π.

4.3 Acceleration and transition risk

Without policy — Suppose that the baseline technology path $A_{(\cdot)}$ is continuously differentiable, with a positive derivative. Let $\hat{A} \equiv \lim_{t\to\infty} A_t$ be finite or infinite.

As implied above, fixing policy, whether acceleration increases or decreases cumulative risk depends on whether ζ is greater or less than 1. This can, again, be seen by integrating the hazard curve with respect to A:

$$X = \int_0^\infty \bar{\delta} A_t^\alpha \dot{A}_t^\zeta dt = \int_{A_0}^{\hat{A}} \bar{\delta} A^\alpha \dot{A}_A^{\zeta-1} dA.$$

Given acceleration $\tilde{A}_{(\cdot)}$ from $\underline{A} \in [A_0, \hat{A})$ to $\overline{A} \in (\underline{A}, \hat{A}]$, cumulative risk equals

$$\tilde{X} = X + \int_{\underline{A}}^{\overline{A}} \bar{\delta} A^\alpha \big(\dot{\tilde{A}}^{\zeta-1} - \dot{A}_A^{\zeta-1}\big) dA.$$

The integral is negative if $\zeta < 1$, zero if $\zeta = 1$, and positive if $\zeta > 1$.

Because the Jones models implicitly adopt $\zeta = 1$, though there is a technology level $\hat{A} < \infty$ at which it is optimal to halt technological development (Appendix A.5), the speed of technological development before $\hat{A}$ does not affect cumulative risk.

With policy — We have seen that the impact of acceleration on cumulative transition risk is ambiguous absent policy. Under optimal policy it remains ambiguous, but a tendency for acceleration to lower cumulative risk may be reintroduced.

For simplicity and focus, we will assume that A grows at a constant rate g and that $\gamma > 1$. Also, since given exponential growth $g_{\dot{A}} = g$, we will impose

$$\beta > \alpha + \zeta, \tag{28}$$

which, rather than $\beta > \alpha$, is now the condition necessary for survival without $C_t = A_t x_t \to 0$. Under these conditions, since $\dot{A}$ is proportional to A, the planner's problem is precisely as described in Section 3.3, with $\alpha + \zeta$ taking the place of α (up to a coefficient g^ζ that can be incorporated into $\bar{\delta}$). Baseline x and δ paths, and S_∞, are unchanged. The existential risk Kuznets curve remains.

Let A^* denote the uppermost technology level at which it is optimal to set $x = 1$ on the baseline technology path. Since the first-order condition

$$\frac{\partial u}{\partial x_t}(A_t, x_t) \geq \frac{\partial \delta}{\partial x_t}(A_t, x_t) v_t \quad \Longrightarrow \quad A_t^{1-\gamma} x_t^{-\gamma} \geq \bar{\delta} A_t^{\alpha} \beta x_t^{\beta-1} v_t$$

must be satisfied everywhere and hold with equality for $x < 1$, we have

$$x_A = \begin{cases} 1 & A \leq A^*, \\ \left(\bar{\delta} \beta A^{\alpha+\gamma-1} \dot{A}_A^{\zeta} v_A \right)^{-\frac{1}{\beta+\gamma-1}} & A > A^*. \end{cases} \tag{29}$$

Substituting (29) into the expression for cumulative risk

$$X = \int_{A_0}^{\infty} \bar{\delta}\, A^{\alpha} \dot{A}_A^{\zeta-1}\, x_A^{\beta}\, dA, \tag{30}$$

we have

$$X = \int_{A_0}^{A^*} \bar{\delta} A^{\alpha} \dot{A}_A^{\zeta-1} dA + \int_{A^*}^{\infty} \left(\bar{\delta}^{1-\gamma} \beta^{\beta} A^{(\beta-\alpha)(\gamma-1)} v_A^{\beta} \right)^{-\frac{1}{\beta+\gamma-1}} \dot{A}_A^{\zeta \frac{\gamma-1}{\beta+\gamma-1} - 1} dA. \tag{31}$$

Recall that a technology path $\tilde{A}_{(\cdot)}$ is an acceleration if $\dot{\tilde{A}}_A > \dot{A}$ for technology levels $\underline{A} \in [A_0, \infty)$ to $\overline{A} \in (\underline{A}, \infty]$. With or without policy, an acceleration affects cumulative risk directly, by changing the technology growth rate from $\underline{A}$ to $\overline{A}$. With policy, an acceleration also affects cumulative risk indirectly by affecting v_A for $A \in$

$[\underline{A}, \overline{A})$, which affects policy at this range of technology levels.

Under hazard function (5), faster technology growth is always preferred, as explained following (12). A future with faster growth is more valuable, so an acceleration from $\underline{A}$ to $\overline{A}$ raises v_A for $A \in [\underline{A}, \overline{A})$. Under hazard function (27), this no longer holds. Unlike an increase to A_t, an increase to $\dot{A}_t$ brings no contemporaneous benefit, but it now imposes risks that can still be mitigated only with less contemporaneous consumption. We can see that growth is sometimes undesirable most plainly when $\alpha = -1$, $\zeta = 1$: again, this is the Russian roulette model, and as Jones finds, with $\gamma > 1$, the planner grows technology only to a finite level.[18]

These complexities are avoided when we focus on instantaneous accelerations. The impact of an acceleration from $\underline{A}$ to $\overline{A}$ on v_A, for $A \in [\underline{A}, \overline{A})$, falls to zero as $\overline{A} - \underline{A} \to 0$. The impact of a *brief* acceleration on cumulative risk is therefore approximately the impact found when we ignore impacts on v_A.

Proposition 3. ***Instantaneous acceleration and transition risk***

Given hazard function (27) and technology path (4), choose a technology level $\underline{A} > 1$ *and growth rate* $\dot{\underline{A}} > \dot{A}_{\underline{A}}$. *If*

a. $\underline{A} \geq A^*$ *and* $\zeta < (=, >)\, 1 + \frac{\beta}{\gamma - 1}$, *or*

b. $\underline{A} < A^*$, $\zeta < (=, >)\, 1$, *and* $\dot{\underline{A}}$ *maintains (29)* $= 1$ *at* $A = \underline{A}$,

then $\Delta_{\underline{A}, \dot{\underline{A}}} < (=, >)\, 0$.

The result follows essentially immediately from the exponent on $\dot{A}_A$ in (30).[19] In

[18] In this more general model, the optimality of *stagnation* is knife-edge (Appendix A.5), but the result that accelerating from $\underline{A}$ to $\overline{A}$ does not necessarily raise v_A for $A \in [\underline{A}, \overline{A})$ is not.

[19] A rigorous proof may be found in Appendix B.3.

particular, instantaneous acceleration after A^* lowers cumulative risk as long as

$$\zeta \frac{\gamma - 1}{\beta + \gamma - 1} - 1 < 0 \implies \zeta < 1 + \frac{\beta}{\gamma - 1}. \tag{32}$$

It is sufficient, though not necessary, for (32) that

$$\zeta \leq 1 \quad \text{or} \quad \alpha \geq -1,\ \gamma \leq 2.$$

The $\zeta \leq 1$ case follows from the fact that $\frac{\gamma-1}{\beta+\gamma-1} < 1$. The $\alpha \geq -1$, $\gamma \leq 2$ case follows from the fact that if $\alpha \geq -1$, then, by (28), $\zeta < \beta + 1$, so $\frac{\zeta}{\beta+1} < 1$. Since macroeconomic estimates of $\gamma \leq 2$ are standard, this result suggests that accelerations lower cumulative risk on the optimal path in the context of transition risk, at least if they occur late enough in time that mitigation is already underway.

Again, this is without considering the fact that an increase to future growth can change the value of the future. Though the direction of this change is ambiguous, it is often taken for granted that, on a conventional discount rate, faster technology growth would currently be a benefit. If so, this is another channel through which a (positive-duration) acceleration increases concern for safety.

It may be counterintuitive that instantaneous acceleration reduces risk only when γ lies *below* a bound. The result is due to the fact that, when γ is high, the marginal utility of consumption rises rapidly as x is cut, so following an acceleration, a small cut to x suffices to equalize the marginal utility of consumption with that of safety spending. The higher γ is, the more quickly x falls as A rises, but the *less* sensitive x is to a change in $\partial\delta/\partial x$—e.g. an increase due to higher $\dot{A}$—at a given value of A.

4.4 Discussion

Nonrivalry in safety effort — Hazard function (27) is explored here mainly for its simplicity and similarity to (5). This functional form overemphasizes a channel through which the risks posed by a series of technological developments can be cheaper to mitigate if they occur at once than if they occur in sequence. Suppose that $\beta \approx 1$, that $\zeta = 1$, and that two small A-increases—"experiments"—can occur in sequence or simultaneously. If they occur in sequence, halving the risk posed by each requires halving x and thus consumption for two periods in a row. If they occur simultaneously, the same risk-reduction only requires halving consumption for one period.

For some kinds of experiments and safety efforts, this "nonrivalry" assumption is reasonable. Monitoring wastewater to detect pandemics early reduces the risk posed by the relevant biological experiments by a proportion independent of how many are underway. For other kinds, the assumption is not reasonable: e.g. it does not apply to the safety equipment that must be used at each lab (c.f. Appendix A.4.2).

This model is thus in no sense a thorough study of the relationship between growth and transition risk. It is intended only to offer two limited lessons. First, under optimal policy, the effect of acceleration on transition risk remains ambiguous. Second, the presence of an optimal policy response can change the conditions under which acceleration lowers risk, and can relax them to the extent that safety efforts are nonrival across contemporaneous risks.

Stagnation vs. deceleration — When $\zeta > 0$, complete stagnation ($\dot{A} = 0$) is always the safest path of all. Nevertheless, we have seen with and without policy that given a positive growth rate, faster growth can decrease risk.

This is because, given stagnation at $\hat{A}$, levels $A > \hat{A}$ are never attained. Cumulative risk is thus not (30) but (30) with the ∞ replaced by $\hat{A}$. Absent stagnation, all levels of A are attained; the growth rate only determines the risk endured at each. The direct cost of faster progress during a given A-range (higher risk per unit time) is partially, and may be more than fully, outweighed by the fact that faster progress motivates more mitigation at each point in time, in combination with the familiar fact that when progress is faster we do not linger in a given A-range as long.

5 Conclusion

Human activity can create or mitigate existential risks. The framework presented here illustrates that, under various conditions, existential risk should be expected to exhibit a Kuznets curve. This observation offers a potential economic explanation for the claim that we are in a "time of perils". We may be advanced enough to be able to destroy ourselves, but not yet rich enough that we are willing to make large sacrifices for the sake of safety. If we are indeed living through the time of perils, reductions to existential risk today have massive long-term consequences.

This framework also highlights a channel through which some efforts intended to reduce existential risk may backfire. In the absence of policy, when risk is posed by the existence of advanced technologies, broad-based decelerations to technological development typically worsen or do not affect the odds of long-term survival. Given an optimal policy response, even by a policymaker with little concern for the long-term future, this impact is magnified. The impact can be significant, with proportional consumption decreases having comparable impacts to proportional increases in the planner's rate of time preference. In the extreme, permanent stagnation can make a catastrophe inevitable that might otherwise have been avoided.

This lesson comes with three caveats. First, it is not an argument against regulating the use of risky technologies. Indeed, a primary channel explored here through which technological development lowers risk is that it hastens the day when regulation is severe. Some recent reactions to calls for heavy AI regulation, e.g. that of Andreessen (2023), might be read as expressing the view that our "x" should never be set far below one. If that is so, it is not for reasons presented in this paper.

Second, when risk is posed by the *development* of advanced technologies, the effect of acceleration on risk is ambiguous. In the "transition risk" models of Jones (2016, 2024), acceleration does not affect cumulative risk. Under slight modifications to these models, the impact may be positive or negative. Policy may facilitate a tendency for acceleration to weakly decrease cumulative risk, as illustrated in Section 4; but it seems likely that in other plausible models it would not.

Third, where we have found that policy magnifies a negative link between acceleration and cumulative risk, we have assumed that policy is optimal. If it is not, then the impact of acceleration on cumulative risk may be reduced or even overturned, as illustrated in Section 3.1.[20] The appropriate lesson about the impact of policy on the relationship between acceleration and risk is only that, to the extent that the policy regime equates or will eventually equate the marginal utility of consumption to the marginal expected discounted utility of safety expenditure, consumption-increasing technological development has the unseen benefit of speeding future safety efforts. For slowing technological development to lower cumulative risk, there must be a policy failure severe and lasting enough to outweigh this potentially large benefit.

In this light, further research on the nature of policy distortions around the regulation of risky technologies would be valuable. Exploring the long-term implications

[20]Shulman and Thornley (2024) argue that the policy response to existential risk is in fact far from optimal, even under a conventional discount rate.

of other models of anthropogenic existential risk, and of optimal policy in the face of it, could be valuable as well, to better characterize the scope of the result that optimally regulated acceleration weakly lowers cumulative risk. If plausible models are found under which the result is overturned, this will naturally pose important questions which can only be answered empirically. For now, however, the results presented here suggest that even those exclusively concerned with reducing cumulative existential risk should often cheer technological advances despite their short-term hazards, and advocate risk-reduction measures today only when they are sufficiently targeted and the costs to broad-based technological progress are sufficiently small.

Appendices

A Supplemental materials

A.1 Calibrating the elasticity of the hazard rate to safety expenditures

Shulman and Thornley (2024) estimate that well-targeted expenditures of $400B over the next decade would reduce the probability of existential catastrophe over the next decade by at least 0.1% in absolute terms, from a baseline of 1.85%.

The scale of the magnitude of the risk is taken from Ord's (2020, p. 167) educated guesses and may be disputed. However, an estimate of β depends only on the *proportion* by which a given consumption sacrifice will reduce the hazard rate. We will rely on Shulman and Thornley's assessment that expenditures of $400B would

multiply the probability of existential catastrophe over the next decade by at most

$$1 - \frac{0.1\%}{1.85\%} \approx 0.946, \tag{33}$$

while remaining agnostic about the the magnitude of the probability. For instance, we are trusting their assessments of the extent to which disease monitoring expenditures would be able to prevent existentially hazardous anthropogenic pandemics by helping authorities to contain them early, while remaining agnostic about the probability per year that such a pandemic will arise.

Global consumption per year is currently approximately \$72.5T.[21] If real consumption grows at 2% per year and the relevant interest rate is 5% per year, the present value of global consumption over the next ten years is approximately $\$72.5\text{T} \times (1 - e^{-10(0.05-0.02)})/(0.05 - 0.02) \approx \$626.4\text{T}$. A sacrifice of $\$400\text{B} = \$0.4\text{T}$ in today's dollars over the next decade is thus a sacrifice that multiplies consumption by a fraction of

$$1 - \frac{0.4}{626.4} \approx 0.99936. \tag{34}$$

Given $x^\beta < 0.946$ at $x \approx 0.99936$, it follows that

$$\beta > \frac{\log(0.946)}{\log(0.99936)} \approx 86.7.$$

This exercise of course tells us nothing about whether it is reasonable to assume a constant-elasticity hazard function in general. If the Shulman and Thornley estimate is correct within three orders of magnitude, however, it does prove that the hazard

[21] World Bank national accounts data and OECD National Accounts data files: Final consumption expenditure (current US$). Retrieved from https://data.worldbank.org/indicator/NE.CON.TOTL.CD?end=2022&start=2022, May 20, 2024.

function is currently convex over at least some range of feasible consumption levels. This follows immediately from the facts that (34) > (33) and that the hazard rate cannot be cut by a proportion greater than one.

A.2 State risk with policy: growth vs. patience

In the context of the model of Section 3, we will compare the effects on cumulative risk of a sharp and permanent level effect at t, in which A is multiplied by m slightly greater than 1, with the effects of permanently dividing ρ by m. If $\gamma \leq 1$, the similarity of the two interventions is trivial: cumulative risk is infinite both before and after each intervention. We will thus assume $\gamma > 1$.

A sharp and permanent level effect at t, whereby A is multiplied by m slightly greater than 1, amounts to a leap forward of approximately m/g years. This decreases cumulative risk by approximately $\delta_t m/g$.

Before t^*, therefore, the impact of a level effect on cumulative risk rises exponentially with δ_t. Early in time δ_t may be arbitrarily low, so the impact of the level effect on cumulative risk may as well. The impact of a decrease to ρ on cumulative risk, on the other hand, does not change over time before t^*. A decrease to ρ does not affect the hazard rate immediately, but decreases it in the future by pulling forward the regime-change time and changing the path of the hazard rate afterward. These impacts do not depend on *when* (before t^*) ρ is lowered.

By contrast, consider what happens as $v_t \to \bar{v}$. By (13), in the limit,

$$x_t \approx \left(\bar{\delta}\beta\bar{v}\right)^{-\frac{1}{\beta+\gamma-1}} A_t^{-\frac{\alpha+\gamma-1}{\beta+\gamma-1}}. \tag{35}$$

At large t, permanently multiplying A by $m > 1$ multiplies x_s, at each $s \geq t$, by approximately $m^{-\frac{\alpha+\gamma-1}{\beta+\gamma-1}}$. In conjunction, the increase to A_s and the proportional

decrease to x_s multiply δ_s by $m^{-\frac{(\beta-\alpha)(\gamma-1)}{\beta+\gamma-1}}$ for $s \geq t$. Similarly, permanently dividing ρ by $m > 1$ multiplies x_s $(s \geq t)$ by approximately $m^{-\frac{1}{\beta+\gamma-1}}$, which multiplies δ_s $(s \geq t)$ by approximately $m^{-\frac{\beta}{\beta+\gamma-1}}$. The impacts are equal iff

$$\begin{aligned}(\beta-\alpha)(\gamma-1) &= \beta \\ \iff \gamma &= 2 + \frac{\alpha}{\beta-\alpha},\end{aligned} \tag{36}$$

with the level effect more impactful if the left-hand side is greater and the decrease to ρ more impactful if the right-hand side is greater. The growth-based intervention is more impactful when γ is higher, because higher values of γ motivate faster transitions from consumption to risk-reduction.

Since $\beta > \alpha > 0$, expression (36) reveals that the level effect can only be more impactful in this model if $\gamma > 2$. Still, it is notable that mere level effects to growth can ultimately affect the probability of survival at a comparable scale to permanent, equally-proportioned decreases to the social rate of pure time preference (holding technology growth fixed). Put another way, even temporary stagnation can carry long-term costs similar to those of permanently moving ethical attitudes away from concern for the future.

A.3 State risk with policy: generalized results

Sections 3.3–3.4 are set in the environment of Section 3.2. The three components of this environment are the technology path, the function from technology and policy to the hazard rate, and the utility function. A functional form is assumed for each.

Here we will maintain CRRA utility with $\gamma > 1$. We will however greatly relax our assumptions on the technology path and the hazard rate, to identify the conditions

under which the lessons of Sections 3.3–3.4 are maintained.

In Sections A.3.2–A.3.3, generalizing Proposition 1 from Section 3.3, we find that growth motivates increasing concern for safety: it is often optimal to set $x = 1$ early in time and $x \to 0$ late in time. A central result is that, unless lowering risk is so difficult that it is not achieved even with *stagnation in consumption*, the hazard rate is also driven to 0.

In Section A.3.4, generalizing Section 3.4, we likewise find that when a hazard function is compatible with survival, faster technology growth generally increases the probability of survival. The results support the robustness of the lessons drawn from hazard function (5): that survival is likely possible on the optimal path, and that faster consumption technology growth, if optimally regulated, will raise its probability.

A.3.1 Assumptions

Assumptions on technology growth — We will assume throughout only that the technology path $A_{(\cdot)}$ satisfies some or all of the following conditions:

A1. continuous differentiability almost everywhere and right-continuity of the right derivative $\dot{A}$ everywhere, with $\dot{A}_t > 0$ for all t;

A2. $A_0 > 1$;

A3. $\lim_{t \to -\infty} A_t = 0$; and

A4. $\lim_{t \to \infty} A_t = \infty$.

We will call a technology path $A_{(\cdot)}$ admissible if it satisfies A1–A4.

Assumptions on the hazard rate — We will also consider a wider class of hazard functions. Among these, we will find relatively simple conditions under which a given hazard function and a given technology growth path are compatible with survival on the planner's policy.

Return to the three desiderata preceding the introduction of hazard function (5). We will assume weakenings of two of these desiderata directly, and certain results will require a weakening of the third. In particular, we will assume that the hazard rate increases in x no less quickly than in A and is weakly convex in x. For certain results we will assume that the hazard rate does not decrease too quickly in A.

We will add to these the preliminary conditions that $\delta(\cdot)$ is continuously differentiable; that, when consumption equals zero, so that the entire productive capacity of society is dedicated to existential risk reduction, $\delta = 0$; and that otherwise $\delta > 0$.[22]

Formally, we will assume at most that the hazard rate is a function of $A > 0$ and $x \in (0, 1]$ satisfying the following conditions:

D1. $\delta(A, x) > 0$,

D2. $\lim_{x\to 0} \delta(A, x) = \lim_{A\to 0} \delta(A, x) = 0$,

D3. twice continuous differentiability,[23]

D4. $\eta_x(A, x) \geq \eta_A(A, x)$, and

D5. weak concavity in x,

where η_y denotes the elasticity of δ with respect to $y \in \{A, x\}$. We will call a hazard function admissible if it satisfies D1–D5.

[22] Recall that the hazard rate denotes the flow probability of *anthropogenic* existential catastrophe.

[23] We will define $\frac{\partial \delta}{\partial y}(A, 1) \equiv \lim_{x\to 1} \frac{\partial \delta}{\partial y}(A, x)$ for $y \in \{A, x\}$, and allow these derivatives to be infinite.

Note that the constant elasticity hazard function of Sections 3.2–3.4 is admissible, with $\eta_A = \alpha$ and $\eta_x = \beta$ independent of A and x. Note also that we do not require $\eta_A(A, x)$ always to be positive: we allow new technologies to lower the hazard rate at a given degree of foregone consumption.

A.3.2 The end of consumption growth

Let $C^* \equiv \lim_{t\to\infty} A_t x_t$, when this limit is defined.

Given hazard function (5), $C^* = \infty$, by (19) from Proposition 1. However, some admissible hazard functions motivate decreases to x fast enough that we do not have $C^* = \infty$. C^* may be finite, or C_t may oscillate indefinitely.

Proposition 4. ***The end of consumption growth***

Given an admissible hazard function $\delta(\cdot)$, define

$$R(C) \equiv \lim_{A\to\infty} \frac{\partial\delta}{\partial x}\Big(A, \frac{C}{A}\Big)\frac{C^\gamma}{A}\bar{v}, \tag{37}$$
$$R^* \equiv \lim_{C\to\infty} R(C).$$

Given an admissible technology path and hazard function,

a) *If $R^* \leq 1$, then $C^* = \infty$.*

b) *If $R^* > 1$, then $C^* \neq \infty$.*

Proof. See Appendix B.4. □

To interpret the result, recall that $x = C/A$. (37) characterizes, if C is fixed even as A grows, what happens to the ratio of the marginal value of lowering x via increased safety ($\frac{\partial\delta}{\partial x} \cdot v$) to the marginal utility of raising x via increased consumption ($AC^{-\gamma}$). If the ratio approaches 1, then it is optimal for consumption to stagnate in the

long run at C. If the ratio is greater than 1 for sufficiently large C, therefore, then stagnation at some finite C is optimal.

Recall from (17) that $\bar{v} \equiv \frac{1}{\rho(\gamma-1)}$. When $R(C) > 0$, therefore, $R(C)$ decreases in ρ. A lower discount rate can thus shift R^* from below to above 1, resulting in stagnation when there would otherwise have been long-run consumption growth, but never the reverse. Consumption stagnation is not in general desirable when ρ is sufficiently low, or undesirable when ρ is sufficiently large: for many hazard functions, as shown at the end of the next subsection, R^* is above 1 (even infinite) or below 1 (even 0) for any $\rho > 0$. Still, Proposition 4 illustrates how calls for an "end to growth" may be compatible with this model. Concern for the future can motivate controls on technological deployment strict enough to halt growth in *consumption*, despite the tendency for accelerating *technological development* to lower cumulative risk.

A.3.3 The Kuznets curve generalized

Proposition 5. ***The Kuznets curve generalized***

Given an admissible technology path and hazard function,

a) $\lim_{t\to-\infty} x_t = 1$.

If η_A *is bounded above* $1-\gamma$*, then* $\lim_{t\to\infty} x_t = 0$.

b) $\lim_{t\to-\infty} \delta_t = 0$.

If $C^* = \infty$*, then* $\lim_{t\to\infty} \delta_t = 0$.

If $C^* \neq \infty$*,* η_A *is bounded above* $1-\gamma$*, and* η_x *is upper-bounded, then* $\lim_{t\to\infty} \delta_t \neq 0$.

Proof. The proof of (a) is given in Appendix B.5. The proof of (b) is as follows.

By D1, D2, and D5, $\delta(A, x)$ is non-decreasing in x. So for all t, $\delta_t \leq \delta(A_t, 1)$. By D2, $\lim_{A\to 0} \delta(A, 1) = 0$. So by A3, $\lim_{t\to-\infty} \delta_t = 0$.

For the positive limit, begin with the weak first-order condition that the marginal flow utility of increasing x must weakly exceed the marginal cost via an increased hazard rate. Then multiply both sides by x_t:

$$\begin{aligned} A_t^{1-\gamma} x_t^{-\gamma} &\geq \frac{\partial \delta}{\partial x}(A_t, x_t)\, v_t \\ \implies (A_t x_t)^{1-\gamma} &\geq \frac{\partial \delta}{\partial x}(A_t, x_t) x_t\, v_t. \end{aligned} \tag{38}$$

If $C^* = \infty$, the left-hand side of (38) tends to 0. Since v is (eventually) positive and does not fall by D4, $\frac{\partial \delta}{\partial x} x \to 0$. Since $\frac{\partial \delta}{\partial x} x \geq \delta$ by D1 and D5, $\delta \to 0$.

If η_A is bounded above $1 - \gamma$, $\lim_{t\to\infty} x_t = 0$ by (a). Since eventually $x_t < 1$, eventually (38) holds with equality. If $C^* \neq \infty$, the left-hand side does not tend to 0 in the limit. Because v_t is upper-bounded, $\frac{\partial \delta}{\partial x} x$ does not tend to zero either. So if $\eta_x \equiv \frac{\partial \delta}{\partial x} \frac{x}{\delta}$ is upper-bounded, $\delta \not\to 0$. □

Part (b) of the proposition stems from the fact that, as long as consumption rises without bound, its marginal utility falls to zero. If the hazard rate does not also fall to zero, the marginal value of sacrificing consumption to lower it further stays positive. The hazard rate must therefore fall to zero.

Even so, unbounded consumption growth does not necessarily coincide with a positive probability of survival. To achieve $S_\infty > 0$, δ_t must not only fall to 0 but fall sufficiently quickly. This in turn is guaranteed whenever consumption rises sufficiently quickly, which holds under a strengthening of the condition for unbounded consumption growth from Proposition 4.

Proposition 6. ***Survival generalized***

Given an admissible hazard function $\delta(\cdot)$ and an admissible technology path $A_{(\cdot)}$ such

that, for some $k > 1$ and some $\underline{t}$ we have

$$A_t \geq t^{\frac{k}{\gamma-1}} \quad \forall t > \underline{t}, \tag{39}$$

define

$$\tilde{R}(k) \equiv \lim_{t\to\infty} \frac{\partial\delta}{\partial x}\Big(A_t, \frac{t^{\frac{k}{\gamma-1}}}{A_t}\Big)\frac{t^{\frac{k\gamma}{\gamma-1}}}{A_t}\bar{v}.$$

a) *If $\lim_{k\to 1^+} \tilde{R}(k) < 1$, then $\exists \underline{t} \,:\, C_t > t^{\frac{1}{\gamma-1}} \quad \forall t > \underline{t}$ and $S_\infty > 0$.*

b) *If $\lim_{k\to 1^-} \tilde{R}(k) > 1$, then $\exists \underline{t} \,:\, C_t < t^{\frac{1}{\gamma-1}} \quad \forall t > \underline{t}$.*

If in addition η_x is upper-bounded, then $S_\infty = 0$.

Proof. See Appendix B.6. □

Observe that, similar to $R(\cdot)$, $\tilde{R}(k)$ is the long-run ratio of the marginal value of lowering risk to the marginal value of increasing consumption when

$$C_t \propto t^{\frac{k}{\gamma-1}}. \tag{40}$$

If $\tilde{R}(k) < 1$ on this consumption path, for some $k > 1$, then on this path consumption grows too slowly. It is eventually preferable to raise x_t above its implied level of approximately $t^{\frac{k}{\gamma-1}}/A_t$. So if $\lim_{k\to 1^+} \tilde{R}(k) < 1$, C_t eventually grows more quickly than (40) for some $k > 1$ on the optimal path. Conversely, if $\lim_{k\to 1^-} \tilde{R}(k) \geq 1$, C_t eventually grows more slowly than (40) for $k = 1$.

If C_t grows more quickly than (40) for some $k > 1$, then the left-hand side of (38) falls more quickly than t^{-k} for some $k > 1$. So $\frac{\partial\delta}{\partial x}x$ does as well. Recalling that $\delta < \frac{\partial\delta}{\partial x}x$, this ensures a positive probability of survival.

If C_t grows more slowly than (40) for $k = 1$, then the left-hand side of (38) falls more slowly than $1/t$. The right-hand side equals $\frac{\partial \delta}{\partial x} x \cdot v = \eta_x / \delta \cdot v$. If η_x is upper-bounded, δ falls more slowly than $1/t$. Cumulative risk is therefore infinite, and survival is impossible.

For illustration, let us evaluate the constant elasticity hazard function of Section 3.2 for the case of exponential growth at rate g.

$$\begin{aligned} \tilde{R}(k) &= \lim_{t \to \infty} \bar{\delta} e^{\alpha g t} \beta \Big(\frac{t^{\frac{k}{\gamma - 1}}}{e^{gt}} \Big)^{\beta - 1} \frac{t^{\frac{k\gamma}{\gamma - 1}}}{e^{gt}} \bar{v} \\ &= \bar{\delta} \beta \bar{v} \lim_{t \to \infty} e^{-(\beta - \alpha) g t} t^{\frac{\beta + \gamma - 1}{\gamma - 1} k} = 0 \end{aligned} \tag{41}$$

for any k, since $\beta > \alpha$. So $\lim_{k \to 1^+} \tilde{R}(k) = 0 < 1$. Part (a) of Proposition 6 thus generalizes the conclusion of (26) that, with hazard function (5), consumption grows at least as quickly as a sufficient power function (in fact it grows exponentially) and that there is a positive probability of survival.

By contrast, consider the constant elasticity hazard function but with $\alpha = \beta$. In this case, (41) $= \infty$ for any k, so $\lim_{k \to 1^-} \tilde{R}(k) = \infty > 1$. Also, η_x is constant at β, and so upper-bounded. $\delta(A, x) = Ax$ is thus an example of a hazard function satisfying D1–D5 for which the probability of survival on the optimal path is zero given exponential technology growth (and indeed given any $A_{(\cdot)}$ that is eventually bounded above zero).

A.3.4 Acceleration and state risk generalized

For any admissible hazard function, the lessons of Section 3.4 are essentially maintained. The effect of a temporary level effect on the probability of survival is ambiguous. However, if the probability of survival is positive on the planner-optimal policy

path, given the baseline technology path, then an acceleration to technological development increases the probability of survival. If the probability of survival is zero on the planner-optimal policy path, then an acceleration to technological development may increase the probability of survival or have no effect.

Proposition 7. ***Acceleration and state risk generalized***

Choose an admissible technology path $A_{(.)}$ *and hazard function* $\delta(\cdot)$.
Given $\underline{A}$, $\dot{\tilde{A}}$ *with* $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$,

a) $\Delta_{\underline{A},\dot{\tilde{A}}} = \delta_{\underline{A}}(\dot{\tilde{A}}^{-1} - \dot{A}_{\underline{A}}^{-1}) < 0.$

Given an acceleration $\tilde{A}_{(.)}$ *from* $\underline{A}$ *to* $\overline{A}$,

b) *If* $X < \infty$, *then* $\tilde{X} \leq X + \int_{\underline{A}}^{\overline{A}} \Delta_{A,\dot{\tilde{A}}_A} dA < X$.

c) *If* $X = \infty$ *and the acceleration is temporary, then* $\tilde{X} = \infty$.
If $X = \infty$ *and the acceleration is permanent, then* $\tilde{X}$ *may be finite or infinite.*

Proof. See Appendix B.7. □

The intuition is the same as illustrated in Section 3.4. Acceleration in effect horizontally rescales all or part of the hazard curve by leaving less time spent at each state. It may also induce more stringent policy at each state, in which case the weak inequality of part (b) is strict.

A.3.5 Discussion

Accelerations vs. level effects — Given a technology path $A_{(.)}$ satisfying A1 and A4, say that a differentiable technology path $\tilde{A}_{(.)}$ is a level effect to $A_{(.)}$ (at time 0) if

$$\exists m > 1 \ : \ \tilde{A}_t = mA_t \ \ \forall t.$$

When technology growth is exponential, level effects are (sharp) temporary accelerations. Otherwise, they may be distinct.

Unlike temporary accelerations, level effects do not always decrease cumulative risk outside the exponential growth context. Consider for example hazard function (5) with a technology path $A_{(\cdot)}$ that is nearly stagnant for an arbitrarily long period, say for $t \leq 99$; that grows exponentially at $t > 99$; and for which the implied regime-change time is $t^* = 100$. A level effect—a jump in the technology level at $t = 0$—then raises the technology level during the arbitrarily long period of stagnation, which non-negligibly raises cumulative risk, while lowering cumulative risk only negligibly by cutting a vertical slice from the hazard curve following $t = 99$.

The direction of technical change — This is a model in which there is a single dimension to technological development. Inventions simply occur in sequence, each of which increases potential consumption and has some effect on the hazard rate at any given level of consumption. In practice, however, technological development is surely at least somewhat *directed*: tradeoffs between consumption and risk in later periods are affected by the extent to which policymakers and firms in earlier periods have developed various types of technology. Consider for example the "richer model" of Jones (2016), in which increases in the value of life relative to consumption motivate increases not only in health spending but also in medical R&D.

In positing a baseline sequence of maximum potential consumption levels $\{A_t\}$ and a hazard function $\delta(A, \cdot)$, we are simply describing a path of possibilities frontiers over time, not embedding any assumptions about how this path is generated. In particular, we are not assuming that there is only one way it is possible for technology to unfold. If we posit a wider space of possible production technologies than the sequence adopted on the baseline path, we must simply

clarify that our results only pertain to "accelerations" in the sense of increases to the rate of motion along the baseline path. Subsidizing the development of risky technologies that would not otherwise have been invented, or choosing a technology path on which they are invented sooner than they would have been but risk-decreasing technologies are not, does not necessarily lower cumulative risk.[24]

In the next section (Appendix A.4), the lessons of this generalized model are used to explore two particular hazard functions that may be of interest. The first illustrates that, early in time, the hazard rate may increase alongside smooth declines in x. The second is "microfounded" by an assumption that increases in safety expenditure lower risk through redundant safeguards.

A.4 State risk with policy: Two more hazard functions of interest

We will assume that technology grows at a constant rate $g > 0$.

[24]In addition to modeling the policy choice about how much consumption to sacrifice for an instantaneous reduction to the hazard rate, an earlier version of this paper models the technology path as directed by policy as well. The growth model is semi-endogenous, so total potential technology growth is driven by exogenous population growth, but research is optimally allocated between risk-increasing "consumption technology" and risk-decreasing "safety technology". Conceptually, that model sheds light on the same question as this one—how acceleration affects cumulative risk, given an endogenous policy response—but the objects of study are accelerations to population rather than to technology itself. Numerical estimation suggests that acceleration weakly decreases cumulative risk in that context as well, for the same reasons as it does here. When population growth is accelerated, and labor is allocated optimally across fields, civilization traverses roughly the same technology path but more quickly. When future population growth is anticipated to be faster, the value of the future is higher (due to faster future technological development even if larger populations are not valued more intrinsically), so optimal policy shifts the technology path in a safer direction.

A.4.1 A lower Inada condition on safety

As shown in Section 3.3, given a constant elasticity hazard function, δ rises as long as it remains optimal to maximize consumption, and falls immediately once it becomes optimal to begin choosing sub-maximal consumption out of concern for safety. This result is arguably at odds with the experience of the last century, during which the hazard rate has arguably risen while existential safety expenditures have risen (from essentially 0). We will therefore here explore how to tweak the hazard function so that the Kuznets curve is smoothed, and the policy choice variable falls even early in time while the hazard rate is still rising.

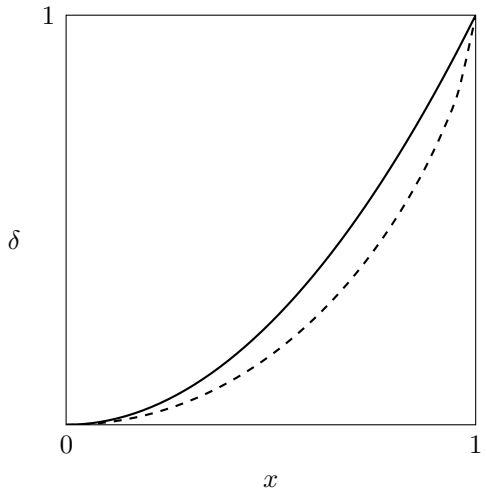
A constant elasticity hazard function generates a distinct pair of regimes for the same reason here as in Stokey (1998): because, when $x = 1$, marginal "safety expenditures"—decreases to x—produce only finite marginal benefits. That is, there is no "lower Inada condition on safety". We will say that a hazard function exhibits a lower Inada condition on safety if $\lim_{x\to 1} \frac{\partial \delta}{\partial x} = \infty$. Under this condition, it is optimal to set $x_t < 1$ as long as $v_t > 0$: as long as civilization is worth preserving at all, some expenditures on existential risk reduction are worthwhile.

Not every hazard function with a lower Inada condition on safety behaves like a smoothed version of a constant elasticity hazard function. If the inverse of the hazard function is too concave around $x = 1$ (when A is low), then x may fall rapidly, rather than mildly, from the outset, yielding no early period during which $x \approx 1$. If it is not concave enough around $x = 1$, on the other hand, then early decreases to x produce significant decreases to δ, so that the hazard rate falls even early in time.

One class of hazard functions with the desired features is

$$\delta_t = \bar{\delta} A_t^{\alpha} x_t^{\beta} \frac{1 - (1 - x_t)^{\epsilon}}{x_t}, \quad \epsilon \in \left(\frac{1}{2}, 1\right), \tag{42}$$

where the conditions on parameters other than ϵ are as in (5). The distinction between the hazard functions is illustrated below for the case of $\bar{\delta}A^{\alpha} = 1$, $\epsilon = 0.6$, $\beta = 2$. The solid curve represents the old hazard function; the dashed curve represents the new hazard function, vertical at $x = 1$.



Note that

$$\lim_{x \to 0} \frac{1 - (1 - x_t)^{\epsilon}}{x_t} = \epsilon,$$

so the asymptotics in this case are identical to those in the case of a constant elasticity hazard function (except that the hazard rate is multiplied by ϵ). However, the transition dynamics are different. Though it is now optimal to set $x < 1$ as long as $v > 0$, x now falls smoothly and δ smoothly rises and falls. The paths of risk and policy are illustrated below for $\epsilon = 0.6$, $A_0 = 2.03$, and otherwise the same parameter values as in Table 1.[25]

[25] A_0 is raised slightly in order to maintain that the value of a statistical life-year "today" (at $t = 75$) is four times per capita consumption, and the hazard rate is approximately 0.1%, despite the fact that, in this model, consumption and the hazard rate are slightly less than maximal even early in time.

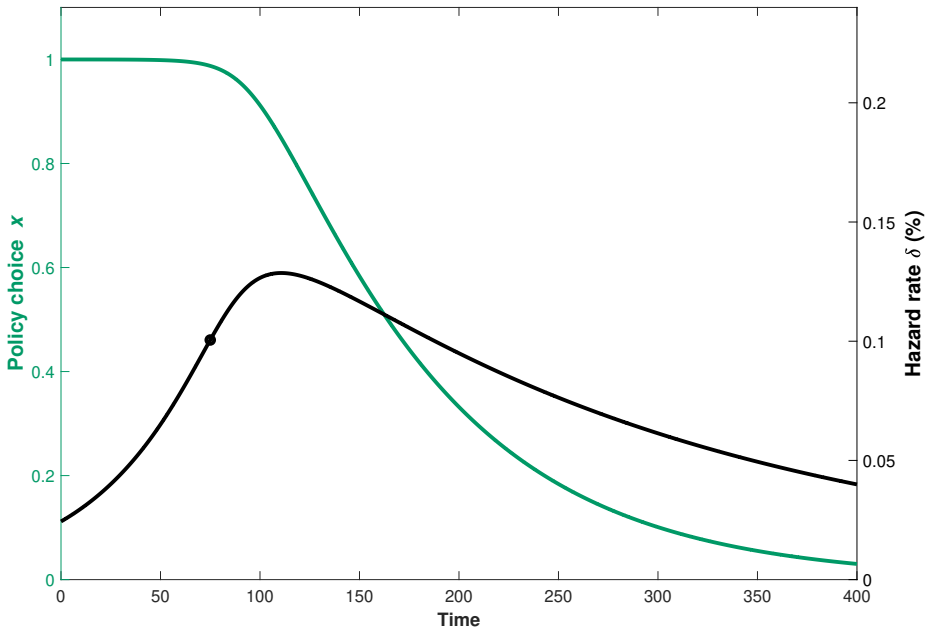


Figure 3: Evolution of the policy choice and the hazard rate along the optimal path given a lower Inada condition on safety expenditure

Derivations and code for replicating the simulation may be found in Appendix C.

A.4.2 Safety in redundancy

The constant elasticity hazard function of Sections 3.2–3.4, and its tweak just above, were chosen for clarity. We might however be interested in a better-founded story about the shape of the hazard function, in which the hazard rate is determined by the production of consumption goods and safety goods. For illustration, one relatively straightforward story would be as follows.

- Each unit of consumption (still produced as $C_t = A_t x_t$) poses some risk p of catastrophe per period in the absence of any safety measures.

- For each unit of the consumption good, if one unit of the safety good (produced

as $H_t = A_t(1 - x_t)$) is allocated to preventing the production process from causing a catastrophe, this fails to prevent a catastrophe with probability $\tilde{b} < 1$. That is, one unit of H per unit of C multiplies the risk posed by each unit of C by $\tilde{b}$, from the baseline of p.

- The probability that the production of a given unit of consumption results in a catastrophe is the probability that (a) there would have been a catastrophe in the absence of any safety measures and (b) all H/C safety measures fail independently: $p\tilde{b}^{H/C}$.

- The probability of survival through a given period is the probability that all C consumption units, independently, do *not* generate a catastrophe: $(1-p\tilde{b}^{H/C})^C$.

In discrete time, the story above would correspond to the hazard function

$$\delta(A_t, x_t) = 1 - \left(1 - p\tilde{b}^{\frac{1-x_t}{x_t}}\right)^{A_t x_t}, \quad \tilde{b} \in (0, 1). \tag{43}$$

The continuous-time analog to (43) is

$$\delta(A_t, x_t) = A_t x_t e^{-b\frac{1-x_t}{x_t}}, \quad b > 0 \tag{44}$$

(see Appendix B.8.1).

Since hazard function (44) lacks any sort of lower Inada condition on $1 - x$, x is fixed at 1, and δ rises, early in time while $v > 0$. After the relevant calculations, Propositions 4–6 tell us that (44) yields a Kuznets curve, with δ eventually falling quickly enough to permit survival.

Proposition 8. ***Long-run policy and risk given safety in redundancy***

Given hazard function (44), the optimal path features

$$\lim_{t\to\infty} x_t t = \frac{b}{g\gamma}, \tag{45}$$

$$\lim_{t\to\infty} g_{\delta t} = -g(\gamma - 1). \tag{46}$$

Proof. See Appendix B.8.2. □

Thus the decline in policy choice here is slower than in the constant elasticity case: x declines proportionally to $1/t$, not exponentially. This is because a redundancy-based model yields a hazard rate that falls rapidly in the policy choice variable: unit decreases in $A_t x_t$, rather than merely proportional increases, generate proportional decreases to δ. In both cases, however, $x_t \to 0$. And in both cases, δ_t declines exponentially, and so quickly enough to permit survival.

Comparing (46) to the limiting expression for g_δ from Proposition 1, we see that, in the limit, the hazard rate declines more quickly in the redundancy-based model than in the original model. This follows from the fact that the extra coefficient on $g(\gamma - 1)$ in the limiting expression for g_δ from Proposition 1 is less than one:

$$\alpha > 0, \gamma > 1 \quad \Longrightarrow \quad \frac{\beta - \alpha}{\beta + \gamma - 1} < 1.$$

Intuitively it is because, in a redundancy-based model, smaller consumption sacrifices (linear rather than proportional) are needed for proportional decreases to the hazard rate. The planner's response to this expanded possibilities frontier comes partially in the form of slower increases in foregone consumption, as described by (45), and partially in the form of faster declines in the hazard rate, as described by (46).

A.5 Transition risk: Optimal technology growth

A.5.1 Without policy, optimality of stagnation given $\zeta = 1$

Suppose first that $\zeta = 1$ and $\alpha = -1$, so that

$$\delta_t = \bar{\delta}\frac{\dot{A}_t}{A_t}.$$

As noted in the body text, this model is precisely the Russian roulette model of Jones (2016), with $\bar{\delta}$ representing the variable there denoted π.

Jones finds in his setting that, with $\gamma > 1$, it is optimal for technology to grow only to a finite level $\hat{A}$. In our notation, this is because stagnation at some $\hat{A}$, with no risk, yields constant flow utility of $u(\hat{A})$ and a constant value of the future of $v(\hat{A}) \equiv u(\hat{A})/\rho$. It is thus optimal to halt growth at the technology level at which the future benefits of stagnating at a slightly higher A equal the costs via temporarily inducing a positive hazard rate:

$$\begin{aligned} v'(\hat{A}) &= \frac{\partial \delta}{\partial \dot{A}} \cdot v(\hat{A}) \\ \implies \frac{u'(\hat{A})}{\rho} &= \frac{\bar{\delta}}{\hat{A}}\frac{u(\hat{A})}{\rho} \qquad (47) \\ \implies \hat{A} &= \Big(\frac{\bar{\delta}+\gamma-1}{\bar{\delta}}\Big)^{\frac{1}{\gamma-1}}. \qquad (48) \end{aligned}$$

When $\alpha = -1$, we can derive an analytic solution for the optimal technology level (48) at which to stagnate. Though this is not possible for other values of α, it is easy to verify that, for any $\alpha \geq -\gamma$, this result does not qualitatively change.

Equality (47) is then modified to

$$\frac{u'(\hat{A})}{\rho} = \bar{\delta}\hat{A}^{\alpha}\frac{u(\hat{A})}{\rho}$$
$$\implies \hat{A}^{-(\alpha+\gamma)} = \bar{\delta}u(\hat{A}). \tag{49}$$

Given $\alpha+\gamma > 0$, the left-hand side falls strictly monotonically from 1 to 0 as $\hat{A}$ rises from 1 to ∞. The right-hand side rises strictly monotonically from 0 to $\bar{\delta}/(\gamma-1) > 0$ as $\hat{A}$ rises from 1 to ∞. There is thus a unique $\hat{A} > 1$ at which (49) is satisfied: that is, at which technology growth is preferred to stagnation iff $A < \hat{A}$.

A.5.2 Without policy, no optimal stagnation given $\zeta \neq 1$

If we further generalize from $\zeta = 1$ to arbitrary ζ, however, we find that the result that stagnation is optimal when $\zeta = 1$ is knife-edge.

Let $v_t(A_{(\cdot)})$ denote the value of the future at $t \geq 0$ given technology path $A_{(\cdot)}$. As baseline, choose a technology path $A_{(\cdot)}$ satisfying A1 and A2.

If $\zeta < 1$, then at every t, and for every technology level $\overline{A} > A_t$, there is a differentiable and weakly increasing technology path $\tilde{A}_{(\cdot)}$ with $\tilde{A}_s = A_s$ for all $s \leq t$, $\tilde{A}_{\bar{t}} = \overline{A}$ for some $\bar{t} > t$, and $v_t(\tilde{A}_{(\cdot)}) > v_t(A_{(\cdot)})$.

To construct such a path, choose t and $\overline{A} > A_t$. Observe that, if $\dot{\tilde{A}}_A$ equals a constant value $\dot{\tilde{A}}$ for $A \in (A_t, \overline{A})$, the cumulative risk endured on path $\tilde{A}_{(\cdot)}$ from A_t to $\overline{A}$ equals

$$\int_{A_t}^{\overline{A}} \bar{\delta}A^{\alpha}\dot{\tilde{A}}^{\zeta-1}dA,$$

which $\to 0$ as $\dot{\tilde{A}} \to \infty$. With $\zeta < 1$, therefore, sufficiently rapid growth from A_t

to $\overline{A}$ approximates an immediate, risk-free jump from A_t to $\overline{A}$, as in the state risk "$\zeta = 0$" case.

Now let

$$\underline{t} \equiv \min\{t : \tilde{A}_t = \overline{A}\} = \frac{\overline{A} - A_t}{\dot{\tilde{A}}},$$
$$\bar{t} \equiv \sup\{t : A_t < \overline{A}\},$$

noting that $\bar{t}$ may be infinite, and choose $\overline{A}$ and $\dot{\tilde{A}}$ so that $\dot{\tilde{A}} > \dot{A}_s$ for all $s \in [t, \underline{t})$. This is possible for some sufficiently high $\dot{\tilde{A}}$ by the right-continuous differentiability of $A_{(\cdot)}$, and ensures that $\tilde{A}_s > A_s$ throughout this interval. Suppose that $\tilde{A}_t = \overline{A}$ for $t \in [\underline{t}, \bar{t}]$ and $\tilde{A}_t = A_t$ for $t > \bar{t}$—i.e. that the new path halts growth at $\overline{A}$ until the old path has caught up, if ever, after which the paths are identical. Then $\tilde{A}_{(\cdot)}$ offers strictly higher consumption than $A_{(\cdot)}$ across $(t, \bar{t})$ in exchange for arbitrary little up-front risk and no subsequent increases in the hazard rate.

Incidentally, this framework makes clear that, in the absence of any costs to technological development besides transitional existential risk, with $\zeta < 1$ there is no optimal continuous technology path. An immediate jump in the technology level is always desirable, and a larger jump is always preferable to a smaller one. Furthermore, if one introduces R&D costs to the model, an optimal path will exist only if the costs are sufficiently convex in the speed of technological development. Otherwise, attempts to identify an optimal technology path will encounter the "chattering" problem: rapid alternations between slow and fast growth will be preferred to continuous growth, because they can achieve a given quantity of technological progress over a given interval of time while contributing less to cumulative risk.

Stagnation is not optimal given $\zeta < 1$ because, due to the "upper Inada condition" on $\delta \propto \dot{A}^{\zeta}$ with $\zeta < 1$, sufficiently fast technological development carries arbitrarily little risk per unit of new technology. Stagnation is not optimal given $\zeta > 1$ because, since $\lim_{\dot{A}\to 0} \frac{\partial \delta}{\partial \dot{A}} = 0$ when $\delta \propto \dot{A}^{\zeta}$ with $\zeta > 1$, sufficiently slow technological development carries arbitrarily little risk per unit of new technology.

To see this, consider the optimal technology growth rate at t given a technology path $A_{(\cdot)}$ with $A_t = \hat{A} > 1$ and $\dot{A}_s = 0$ for $s > t$. Unlike in the $\zeta < 1$ case, there is an optimal technology growth rate to adopt at t: the rate $\dot{A}^*$ that sets the marginal expected utility benefit (via increased future consumption) of marginally increasing $\dot{A}$, per unit time that $\dot{A}$ is increased, equal to the marginal expected utility cost per unit time (via an increased hazard rate at t):

$$v'(\hat{A}) = \bar{\delta}\hat{A}^{\alpha}\zeta\dot{A}^{*\zeta-1}v(\hat{A})$$
$$\implies \dot{A}^* = \left(\frac{\gamma-1}{\bar{\delta}} \cdot \frac{\hat{A}^{-(\alpha+\gamma)}}{1-\hat{A}^{1-\gamma}}\right)^{\frac{1}{\zeta-1}} > 0.$$

Likewise, given a technology path $A_{(\cdot)}$ with $\lim_{t\to\infty} A_t = \hat{A} < \infty$, the optimal technology growth rate must satisfy the equality above in the limit. Since $A_{(\cdot)}$ cannot approach a finite upper asymptote if $\dot{A}$ is bounded above zero, no such technology path is optimal.

A.5.3 With policy, analogous results for ζ-threshold $1 + \frac{\beta}{\gamma-1}$

Throughout this section we will assume hazard function (27) with $\zeta > 0$:

$$\delta_t = \delta(A_t, \dot{A}_t, x_t) = \bar{\delta}A_t^{\alpha}\dot{A}_t^{\zeta}x_t^{\beta} \quad \bar{\delta} > 0,\ \zeta > 0,\ \beta > 1.$$

For simplicity we will also assume that the baseline technology path features stagnation at technology level $\hat{A}$. We will then consider the impact per unit time of an instantaneous marginal increase to the technology growth rate $\dot{A}_t$.

We will see that, in the $\zeta < 1+\frac{\beta}{\gamma-1}$ case, as in the $\zeta < 1$ case without policy, there is no optimal growth rate: sufficiently fast growth is always preferable to stagnation. In the $\zeta > 1 + \frac{\beta}{\gamma-1}$ case, as in the $\zeta > 1$ case without policy, growth may be "too fast", but there is still no technology level at which it is optimal to stagnate.

However, the $\zeta = 1 + \frac{\beta}{\gamma-1}$ case is *not* closely analogous to the $\zeta = 1$ case without policy. Instead, for low values of $\hat{A}$ it resembles the $\zeta < 1 + \frac{\beta}{\gamma-1}$ case, with no optimal technology growth rate, and for high values of $\hat{A}$ it resembles the $\zeta < 1+\frac{\beta}{\gamma-1}$ case, in which slow growth is preferable both to fast growth and to stagnation. Intuitively, this is because $\zeta = 1 + \frac{\beta}{\gamma-1}$ implies $\zeta > 1$. Since slow growth without policy is preferable to stagnation given $\zeta > 1$, and since introducing the option to mitigate risk with $x_t < 1$ does not remove the option of slow growth without policy, introducing the policy option cannot render stagnation optimal.

In this setting, there are two state variables: S_t and A_t. There are two choice variables: policy x_t and the technology growth rate $\dot{A}_t$.[26] Given $S_t = 1$, the marginal net impacts on expected utility of a marginal increase in $\dot{A}_t$, per unit time, is given by the respective derivative of the Hamiltonian expression

$$u(\hat{A}, x_t) - v_t\delta(\hat{A}, \dot{A}_t, x_t) + a_t\dot{A}_t \tag{50}$$

(adapted from Appendix B.1 below), where a is the costate variable on technology.

[26]It would be equivalent, and more standard but in this case more complex, to define a new choice variable ϕ_t such that the technology law of motion is $\dot{A}_t = \phi_t$.

Under the $x_t \leq 1$ constraint, the optimal choice of x_t given $\dot{A}_t$ is given by the first order conditions $\partial\mathcal{L}/\partial x_t = 0$, $\partial\mathcal{L}/\partial\mu_t \geq 0$, $\mu_t\partial\mathcal{L}/\partial\mu_t = 0$ on the Lagrangian

$$\mathcal{L} = u(\hat{A}, x_t) - v_t\delta(\hat{A}, \dot{A}_t, x_t) + a_t\dot{A}_t + \mu(1 - x_t). \tag{51}$$

This reduces to

$$x_t = \min\left(1, \left(\bar{\delta}\beta\hat{A}^{\alpha+\gamma-1}\dot{A}_t^{\zeta}v(\hat{A})\right)^{-\frac{1}{\beta+\gamma-1}}\right), \tag{52}$$

with $\mu_t > 0$ iff the second term of the above minimum—the unconstrained optimal choice of x_t—is greater than 1. (This is adapted from (58)–(59).)

To find the marginal net impact on expected utility of a marginal increase in $\dot{A}_t$ per unit time, given that x_t is set optimally in response, we can take the first derivative of (51) with respect to $\dot{A}_t$ and evaluate it at $x_t =$ (52). Because (50) and (51) are continuously differentiable in $\dot{A}_t$, x_t, and μ_t, by the envelope theorem we can differentiate (51) with respect to $\dot{A}_t$ and then substitute $x_t =$ (52), rather than accounting for the impact of changing $\dot{A}_t$ on the choice of x_t by substituting (52) into (50) and differentiating the result with respect to $\dot{A}_t$.

Finally, given technology level $A_t = \hat{A}$ and permanent stagnation after t, the value of the costate variables at t are straightforward. The value of [saving] civilization at t is $v(\hat{A})$, and the value of a marginal increase in the technology level is the value of an equal marginal increase in consumption at all future periods:

$$v_t = v(\hat{A}) = \frac{1}{\rho} \cdot \frac{\hat{A}^{1-\gamma} - 1}{1 - \gamma},$$
$$a_t = v'(\hat{A}) = \frac{\hat{A}^{-\gamma}}{\rho}.$$

The marginal net impact on expected utility of a marginal increase in $\dot{A}_t$ per unit time is therefore

$$\begin{aligned} d(\dot{A}_t) &\equiv \frac{\hat{A}^{-\gamma}}{\rho} - v(\hat{A})\bar{\delta}\hat{A}^{\alpha}\zeta\dot{A}_t^{\zeta-1}x_t^{\beta} \qquad (53)\\ &= \frac{\hat{A}^{-\gamma}}{\rho} - v(\hat{A})\bar{\delta}\hat{A}^{\alpha}\zeta\dot{A}_t^{\zeta-1}, \qquad \dot{A}_t < \underline{\dot{A}_t};\\ &= \frac{\hat{A}^{-\gamma}}{\rho} - \zeta\Big(\bar{\delta}^{1-\gamma}v(\hat{A})^{1-\gamma}\hat{A}^{(\beta-\alpha)(\gamma-1)}\beta^{\beta}\Big)^{-\frac{1}{\beta+\gamma-1}}\dot{A}_t^{\zeta\frac{\gamma-1}{\beta+\gamma-1}-1}, \quad \dot{A}_t \geq \underline{\dot{A}_t}, \end{aligned}$$

where

$$\underline{\dot{A}_t} \equiv \left(\bar{\delta}\beta\hat{A}^{\alpha+\gamma-1}v(\hat{A})\right)^{-\frac{1}{\zeta}}$$

is the maximum growth rate at which it is optimal to set $x_t = 1$, and $v(\hat{A})$ is as defined above.

If $\zeta < 1 + \frac{\beta}{\gamma-1}$, then the exponent on $\dot{A}_t$ in (53) is negative for $\dot{A}_t \geq \underline{\dot{A}_t}$, so

$$\lim_{\dot{A}_t\to\infty} d(\dot{A}_t) = \hat{A}^{-\gamma}/\rho > 0.$$

As in the $\zeta < 1$ case without policy, this guarantees that sufficiently fast technology growth is always preferable to stagnation.

If $\zeta > 1 + \frac{\beta}{\gamma-1}$, then the exponent on $\dot{A}_t$ in (53) is always positive. There is thus a unique and positive value of $\dot{A}_t$ that sets $d(\dot{A}_t) = 0$, and this is the optimal choice of $\dot{A}_t$. Sufficiently slow technology growth is always preferable to stagnation.

If $\zeta = 1 + \frac{\beta}{\gamma-1}$, then the exponent on $\dot{A}_t$ in (53) is positive for $\dot{A}_t < \underline{\dot{A}_t}$ and zero for $\dot{A}_t \geq \underline{\dot{A}_t}$. So if $d(\underline{\dot{A}_t}) > 0$, there is no optimal growth rate: from the $A_t = \hat{A}$ margin, it is desirable, albeit perhaps briefly, to have technology grow as quickly as possible. If $d(\underline{\dot{A}_t}) < 0$, there is a unique value of $\dot{A}_t$ that sets $d(\dot{A}_t) = 0$, it lies in

$(0, \underline{\dot{A}_t})$, and it is optimal.

Technically, if $d(\underline{\dot{A}_t}) = 0$, then any $\dot{A}_t \geq \underline{\dot{A}_t}$ is optimal at $A_t = \hat{A}$; but once $A_t > \hat{A}$, we will have $d(\underline{\dot{A}_t}) < 0$, and a unique optimal growth rate which is positive but finite.

B Proofs

B.1 Existence and uniqueness of optimal policy path

Necessary and sufficient conditions — The optimization problems analyzed in Sections 3–4, and the supplemental appendices, feature one choice variable x and one state S. Expected flow utility at t is $S_t u(A_t, x_t)$ for a $\mathcal{C}^2$ function $u(\cdot)$, strictly concave in x, with a lower Inada condition on x. The law of motion for S is $-S_t \delta(A_t, \dot{A}_t, x_t)$ for a $\mathcal{C}^2$ function $\delta(\cdot)$. A, $\dot{A}$ are independent of x, so operate as functions of t.

Letting v denote the costate variable on S, the current value Lagrangian corresponding to the problem is

$$\mathcal{L}(S_t, x_t, v_t, \mu_t, t) = S_t u(x_t, t) - v_t S_t \delta(x_t, t) + \mu_t (1 - x_t) \tag{54}$$

(abusing notation by reusing $u(\cdot)$ and $\delta(\cdot)$ as functions of time), where μ_t is the the Lagrange multiplier on x_t. We impose the $x_t \leq 1$ constraint but not the $x_t \geq 0$ because the latter can never bind, by the lower Inada condition on $u(\cdot)$.

(54) satisfies the Mangasarian concavity condition that $\mathcal{L}(\cdot)$ is weakly concave in S and x. So applying Caputo (2005), Theorems 14.3-4 and Lemma 14.1,[27] given continuous paths of $x \in [0, 1]$ and $S \in [0, 1]$ with $S_0 = 1$ and $\dot{S}_t = -S_t \delta(x_t, t)$, we

[27] Caputo (2005) uses the more general present value notation. Because the control problem at hand is exponentially discounted, we here use the simpler current value notation.

have that the x, S path is optimal if—and, given semi-continuity of x and S, only if—for some semi-differentiable path of v and some semi-continuous path of $\mu \geq 0$, at all t the following first-order conditions are satisfied

$$\frac{\partial \mathcal{L}}{\partial x_t}(S_t, x_t, v_t, \mu_t, t) = \mu_t \frac{\partial \mathcal{L}}{\partial \mu_t}(S_t, x_t, v_t, \mu_t, t) = 0, \quad \frac{\partial \mathcal{L}}{\partial \mu_t}(S_t, x_t, v_t, \mu_t, t) \geq 0, \tag{55}$$

as well as the transversality condition that

$$\lim_{t \to \infty} e^{-\rho t} v_t = \lim_{t \to \infty} e^{-\rho t} v_t S_t = 0. \tag{56}$$

Given optimal paths of x and S and corresponding paths of v and μ, v is continuous and satisfies

$$\dot{v}_t = \rho v_t - \frac{\partial \mathcal{L}}{\partial S_t} = \rho v_t - u(x_t, t) - v_t \dot{S}_t = (\rho + \delta(x_t, t)) v_t - u(x_t, t) \tag{57}$$

except at discontinuity points of x, where v's right and left derivatives may differ.

The transversality condition — Given a continuous v path, only

$$x_t = \begin{cases} 1, & \frac{\partial u}{\partial x}(1, t) - \frac{\partial \delta}{\partial x}(1, t) v_t \geq 0; \\ x_t \ : \ \frac{\partial u}{\partial x}(x_t, t) - \frac{\partial \delta}{\partial x}(x_t, t) v_t = 0, & \text{otherwise} \end{cases} \tag{58}$$

$$\mu_t = \frac{\partial u}{\partial x_t}(x_t, t) - \frac{\partial \delta}{\partial x_t}(x_t, t) v_t \tag{59}$$

satisfy (55) for all t. Any such x path is well-defined, by the continuous differentiability of $u(\cdot)$ and $\delta(\cdot)$ in x and the fact that $u(\cdot)$ and $\delta(\cdot)$ strictly increase in x. Any such x path is also right-continuous in time, by

- the twice continuous differentiability of $u(\cdot)$ and $\delta(\cdot)$ (expressed as functions of x, A, and for $\delta(\cdot)$ in Section 4, $\dot{A}$);
- the right-continuity of the right derivative of $A_{(\cdot)}$ in time;
- (for Section 4, given exponential growth or instantaneous acceleration) the right-continuity of the right derivative of $\dot{A}_{(\cdot)}$;

and the implicit function theorem. Any such μ path is then also right-continuous in time by the composition of continuous functions. To show there exists an optimal path, and that only one such path is semi-continuous, it will now suffice to show that there is a unique v path for which (56)–(57) are satisfied given the corresponding x path (58) and its implied S path, and that the corresponding x path is semi-continuous (in fact it is right-continuous).

The solution to differential equation (57) is

$$v_t = e^{\int_0^t (\rho+\delta_s)ds}\left(v_0 - \int_0^t e^{-\int_0^s (\rho+\delta_q)dq} u(x_s, s)ds\right) \tag{60}$$

$$\implies v_0 = \int_0^t e^{-\rho s} S_s u(x_s, s)ds + e^{-\rho t} S_t v_t. \tag{61}$$

Since (61) is continuous in t (by the boundedness of $u(\cdot)$ and the continuous evolution of S) and holds for all t, v satisfies (56)–(57) iff

$$v_0 = \int_0^\infty e^{-\rho t} S_t u(x_t, t)dt. \tag{62}$$

Given (58), v_t determines x_t for all t. Given (57), v_t and x_t determine the right derivative of v for all t. Given v_0, therefore, there is a unique path of v—and thus of x, and thus of S—compatible with (57)–(58). We will now show that there is at least one value of v_0 for which (62) is satisfied, given the corresponding x and S paths. For such a v_0, the corresponding variable paths by construction satisfy (55)–(56).

Existence — Let $v(v_0)$ and $x(v_0)$ denote the unique paths of v and x compatible with (57)–(58) for which $v_0(v_0) = v_0$. By (60), $\lim_{v_0 \to -\infty} v_t(v_0) = -\infty$ for all $t \geq 0$. By (58), therefore, for every $t \geq 0$, there is a $\tilde{v}_0$ such that $x_t(v_0) = 1$ for all $v_0 < \tilde{v}_0$. Let $s \geq 0$ denote a time at which $A_s \geq 1$, and choose $\tilde{v}_0$ low enough that $\tilde{v}_s < 0$ and thus $x_s(\tilde{v}_0) = 1$. By (57), because $u(1, s) \geq 0$, $\dot{\tilde{v}}_t < 0$. We thus have $\tilde{v}_t < 0$, and thus $x_t = 1$, for all $t \geq s$.

Now observe that if $v_0 < \tilde{v}_0$, $v_t(v_0) < v_t(\tilde{v}_0)$ for all t. Otherwise, by the continuity of v with respect to time, there would be a t with $v_t(v_0) = v_t(\tilde{v}_0)$, and integrating (57), with (58) substituted for x_t, would allow us to identify $v_0 = \tilde{v}_0$. Thus, if $v_0 < \tilde{v}_0$, $x_t(v_0) \geq x_t(\tilde{v}_0)$ for all $t \geq 0$. It follows that, for some sufficiently low $\underline{v}_0$, the right-hand side of (62) exceeds the left-hand side.

For every optimization problem under consideration, there is some $\overline{U}$ by which feasible values of the right-hand side of (62) are upper-bounded. So, for $\overline{v}_0 > \overline{U}$, the left-hand side of (62) exceeds the right-hand side.

By (58) and the implicit function theorem, x_t is continuous in v_t for all t. (57) then implies that $\dot{v}_t$ is defined and continuous in v_t for all t, and thus that $v_t(v_0)$, then $x_t(v_0)$, and then the right-hand side of (62) are continuous in v_0 for all t. It follows from the intermediate value theorem that there exists a $v_0 \in (\underline{v}_0, \overline{v}_0)$ for which (62) holds.

Uniqueness — The uniqueness condition of Caputo (2005), Thm. 14.4 does not directly apply because the Lagrangian is linear, not strictly concave, in S. This can be remedied by defining the state variable as e.g. S^2 without affecting any other results.

Uniqueness (among semi-continuous x paths) also follows from the facts that a path is optimal iff v_0 attains its maximum feasible value and that, given (55)–(56),

v_0 determines a unique path for every variable.

B.2 Proof of Proposition 1

Asymptotic constancy of g_v — From (57), because v is the costate on S, it obeys

$$\dot{v}_t = (\rho + \delta_t)v_t - u(C_t) \quad \Longrightarrow \quad g_{vt} = \rho + \delta(A_t, x_t) - \frac{u(A_t x_t)}{v_t}. \tag{63}$$

Let $\tilde{\beta} \equiv \beta + \gamma - 1$. From (13), once x_t is interior we have

$$x_t = A_t^{-\frac{\alpha+\gamma-1}{\tilde{\beta}}} \left(\bar{\delta}\beta v_t\right)^{-\frac{1}{\tilde{\beta}}}. \tag{64}$$

Substituting (64) into (63) yields

$$g_{vt} = g_v(v_t, t) \equiv \begin{cases} \rho + K A_t^{\frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}}} v_t^{-\frac{\beta}{\tilde{\beta}}} + \frac{1}{1-\gamma} v_t^{-1}, & \gamma \neq 1; \\ \rho + \log\left(A_t^{-\frac{\beta-\alpha}{\beta}} \left(\bar{\delta}\beta v_t\right)^{-\frac{1}{\tilde{\beta}}}\right) v_t^{-1}, & \gamma = 1, \end{cases} \tag{65}$$

where $K \equiv \bar{\delta}^{-\frac{1-\gamma}{\tilde{\beta}}} \left(\beta^{-\frac{\beta}{\tilde{\beta}}} - \frac{1}{1-\gamma}\beta^{-\frac{1-\gamma}{\tilde{\beta}}}\right)$.

If $\gamma > 1$, recalling that v_t monotonically increases and that $A_t \to \infty$, the central term of (65) vanishes. Also, in this case, v is upper-bounded, so it approaches an upper bound v^* by the monotone convergence theorem. So $\lim_{t\to\infty} g_{vt}$ is defined, with

$$\lim_{t\to\infty} g_{vt} = \rho + \frac{1}{v^*(1-\gamma)}. \tag{66}$$

This limit cannot be positive, because v is upper-bounded, and it cannot be negative, because v increases with time. So $\lim_{t\to\infty} g_{vt} = 0$, and $v^* = \frac{1}{\rho(\gamma-1)}$.

If $\gamma < 1$, then $K < 0$, and the central term of (65) grows in magnitude without bound, fixing v. v must therefore also grow without bound, or else g_{vt} is eventually negative. Now observe that

$$\begin{aligned}
\dot{g_{vt}} &= KA_t^{\frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}}} v_t^{-\frac{\beta}{\tilde{\beta}}} \Big(\frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}} g - \frac{\beta}{\tilde{\beta}} g_{vt} \Big) - \frac{1/v_t}{1-\gamma} g_{vt} \\
&= \Big(g_{vt} - \rho - \frac{1/v_t}{1-\gamma} \Big) \Big(\frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}} g - \frac{\beta}{\tilde{\beta}} g_{vt} \Big) - \frac{1/v_t}{1-\gamma} g_{vt} \\
&= -\frac{\beta}{\tilde{\beta}} g_{vt}^2 + \Big(\frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}} g + \frac{\beta}{\tilde{\beta}} \rho + \frac{1}{\tilde{\beta} v_t} \Big) g_{vt} - \Big(\rho + \frac{1/v_t}{1-\gamma} \Big) \frac{(\beta-\alpha)(1-\gamma)}{\tilde{\beta}} g.
\end{aligned}$$

This differential equation has two steady states, both positive. Since $1/v_t \to 0$, the quadratic formula tells us that these steady states approach ρ and $g(\beta-\alpha)(1-\gamma)/\beta$, with the former attractive and the latter repulsive. By (7), ρ is higher, and is ruled out as a steady state by the transversality condition (56). Then because the limit

$$\lim_{t\to\infty} \dot{g_v}(g_v, t) > 0 \ \ \forall g_v \in \Big(\frac{(\beta-\alpha)(1-\gamma)}{\beta} g, \rho \Big); \quad < 0 \ \ \forall g_v < \frac{(\beta-\alpha)(1-\gamma)}{\beta} g$$

is defined and continuous in g_v, we must have

$$\lim_{t\to\infty} g_{vt} = \frac{(\beta-\alpha)(1-\gamma)}{\beta} g. \tag{67}$$

Otherwise $g_v \to -\infty$, ruled out by the monotonicity of v, or $g_v \to \rho$, ruled out above.

The $\gamma = 1$ case is analogous to the $\gamma > 1$ case. Differentiating (65) with respect to time yields $\dot{g_{vt}}$ strictly, continuously increasing in g_{vt} from $-\infty$ at $v_t = 0$ to ρ at $v_t = \infty$. There is thus a unique, positive, and repulsive "time-dependent steady state" value of g_v (i.e. g_v for which $\dot{g_v}(g_v, t) = 0$) which declines to zero as $t \to \infty$.

So

$$\lim_{t\to\infty} \dot{g_v}(g_v,t) > 0 \ \ \forall g_v > 0, \quad \lim_{t\to\infty} \dot{g_v}(g_v,t) < 0 \ \ \forall g_v < 0$$

are defined and continuous in g_v, and $g_v \not\to \infty, -\infty$ imply $\lim_{t\to\infty} g_{vt} = 0$.

Asymptotic behavior of other variables — With the asymptotic behavior of g_v pinned down, that of x and C follows immediately, as does that of δ if $\gamma > 1$.

To find the asymptotic behavior of δ given $\gamma \leq 1$, rearrange (65) to get

$$v_t = \frac{u(C_t)}{\rho + \delta_t - g_{vt}}, \tag{68}$$

and substitute (68) into $C_t^{1-\gamma} = \delta_t \beta v_t$ ((11) rearranged) to get

$$\delta_t = \begin{cases} \frac{\rho+\delta_t-g_{vt}}{\beta} \frac{1-\gamma}{1-C_t^{\gamma-1}}, & \gamma < 1; \\ \frac{\rho+\delta_t-g_{vt}}{\beta \log(C_t)}, & \gamma = 1 \end{cases} \implies \delta_t = \begin{cases} \frac{(\rho-g_{vt})(1-\gamma)}{\beta\left(1-C_t^{\gamma-1}\right)-1+\gamma}, & \gamma < 1; \\ \frac{\rho-g_{vt}}{\beta \log(C_t)-1}, & \gamma = 1. \end{cases}$$

If $\gamma < 1$, the limit of g_v (67) and $C \to \infty$ from (22) imply

$$\lim_{t\to\infty} \delta_t = \frac{(\rho - (\beta-\alpha)(1-\gamma)g/\beta)(1-\gamma)}{\beta+\gamma-1}.$$

If $\gamma = 1$, substitute 0 for g_{vt}. By (22), $\exists \underline{C} > 0 \ : \ \lim_{t\to\infty} \frac{C_t}{e^{\frac{\beta-\alpha}{\beta} gt}} = \underline{C}$, so

$$\lim_{t\to\infty} \delta_t t = \lim_{t\to\infty} \frac{\rho - g_{vt}}{\beta\big(\log(C_t/e^{\frac{\beta-\alpha}{\beta}gt}) + \log(e^{\frac{\beta-\alpha}{\beta}gt})\big)/t - 1/t} = \frac{\rho}{(\beta-\alpha)g}.$$

B.3 Proof of Proposition 3

The proof is similar to the proof of Proposition 7a (Appendix B.7.1), which generalizes Proposition 2. As there, $v_{\underline{A}+\epsilon}$ is continuous in ϵ and $\tilde{v}_{\underline{A}+\epsilon}[\epsilon] = v_{\underline{A}+\epsilon}$ for all ϵ. In this setting, however, we cannot assume that $\tilde{v}_A[\epsilon]$ weakly increases in A or that $\tilde{v}_{\underline{A}}[\epsilon] \geq v_{\underline{A}}$ for all ϵ. We will therefore use a different strategy to uniformly bound $\tilde{v}_A[\epsilon]$, for $A \in [\underline{A}, \underline{A}+\epsilon]$, in an interval whose maximum and minimum converge to $v_{\underline{A}}$ as $\epsilon \to 0$.

Let $\underline{t}$ denote the time at which $A_t = \underline{A}$. An acceleration $\tilde{A}_{(\cdot)}$, featuring technology growth rate $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$ until technology level $\underline{A}+\epsilon$, features technology growth at rate $\dot{\tilde{A}}$ across times

$$(\underline{t}, \underline{t} + \epsilon/\dot{\tilde{A}}).$$

More generally, the acceleration path reaches technology level $A \in [\underline{A}, \underline{A}+\epsilon]$ at time

$$\tilde{t}(A) \equiv \underline{t} + (A - \underline{A})/\dot{\tilde{A}}.$$

$\tilde{v}_A[\epsilon]$ is the maximum value of survival $\tilde{v}_{\tilde{t}(A)}$, across feasible policy paths, achievable at $\tilde{t}(A)$ given technology path $\tilde{A}_{(\cdot)}[\epsilon]$. It can thus be lower-bounded by one such achievable value of survival, such as that achieved given $x_t = 1$ for $t \in [\tilde{t}(A), \underline{t}+\epsilon/\dot{\tilde{A}})$. Since $\tilde{A}_t > 1$ throughout this interval, this lower bound is in turn strictly greater than the value of survival at $\tilde{t}(A)$ given *no* flow utility enjoyed throughout the interval.

Remembering that $\tilde{v}_{\underline{A}+\epsilon}[\epsilon] = v_{\underline{A}+\epsilon} > 0$ for any ϵ, we thus have

$$\begin{aligned}\tilde{v}_A[\epsilon] &\geq \int_{\tilde{t}(A)}^{\underline{t}+\epsilon/\dot{\tilde{A}}} e^{-\rho(t-\tilde{t}(A))} e^{-\int_{\tilde{t}(A)}^{t} \bar{\delta}\tilde{A}_s^{\alpha}\dot{\tilde{A}}^{\zeta} ds} u(\tilde{A}_t) dt \\ &\quad + e^{-\rho(\underline{t}+\epsilon/\dot{\tilde{A}}-\tilde{t}(A))} e^{-\int_{\tilde{t}(A)}^{\underline{t}+\epsilon/\dot{\tilde{A}}} \bar{\delta}\tilde{A}_s^{\alpha}\dot{\tilde{A}}^{\zeta} ds} v_{\underline{A}+\epsilon} \\ &> \underline{v}_A[\epsilon] \equiv e^{-\rho(\underline{t}+\epsilon/\dot{\tilde{A}}-\tilde{t}(A))} e^{-\int_{\tilde{t}(A)}^{\underline{t}+\epsilon/\dot{\tilde{A}}} \bar{\delta}\tilde{A}_s^{\alpha}\dot{\tilde{A}}^{\zeta} ds} v_{\underline{A}+\epsilon}. \end{aligned} \tag{69}$$

Because $\tilde{t}(A)$ increases in A, $\underline{v}_A[\epsilon]$ increases in A, so $\underline{v}_A[\epsilon] \geq \underline{v}_{\underline{A}}[\epsilon]$ for all $A \in [\underline{A}, \underline{A}+\epsilon]$.

$\tilde{v}_A[\epsilon]$ can be upper-bounded by the (infeasible) value of survival achieved at $\tilde{t}(A)$ given that, at $t \in [\tilde{t}(A), \underline{t} + \epsilon/\dot{\tilde{A}})$, flow utility equals its supremum of $1/(\gamma-1)$ and the hazard rate equals 0:

$$\begin{aligned}\tilde{v}_A[\epsilon] &< \frac{1}{\gamma-1}\int_{\tilde{t}(A)}^{\underline{t}+\epsilon/\dot{\tilde{A}}} e^{-\rho(t-\tilde{t}(A))} dt + e^{-\rho(\underline{t}+\epsilon/\dot{\tilde{A}}-\tilde{t}(A))} v_{\underline{A}+\epsilon} \\ &< \overline{v}_A[\epsilon] \equiv \frac{1}{\gamma-1}\int_{\tilde{t}(A)}^{\underline{t}+\epsilon/\dot{\tilde{A}}} e^{-\rho(t-\tilde{t}(A))} dt + v_{\underline{A}+\epsilon}. \end{aligned} \tag{70}$$

Because $\tilde{t}(A)$ increases in A, $\underline{v}_A[\epsilon]$ decreases in A, so $\underline{v}_A[\epsilon] \geq \underline{v}_{\underline{A}+\epsilon}[\epsilon]$ for all $A \in [\underline{A}, \underline{A}+\epsilon]$.

From (69), (70), the continuity of $v_{\underline{A}+\epsilon}$ in ϵ, and the fact that $\tilde{v}_{\underline{A}+\epsilon}[\epsilon] = v_{\underline{A}+\epsilon}$ for all ϵ,

$$\lim_{\epsilon\to 0} \underline{v}_{\underline{A}}[\epsilon] = \lim_{\epsilon\to 0} \overline{v}_{\underline{A}+\epsilon}[\epsilon] = v_{\underline{A}}.$$

The proof then proceeds along the lines of the proof of Proposition 7 after (91), with

$$\tilde{x}_A[\epsilon] = \min\left(1, \left(\bar{\delta}\beta A^{\alpha+\gamma-1}\dot{\tilde{A}}^{\zeta}\tilde{v}_A[\epsilon]\right)^{-\frac{1}{\beta+\gamma-1}}\right)$$

in place of (92), ultimately yielding

$$\Delta_{\underline{A},\dot{\tilde{A}}} = \delta(\underline{A}, \dot{\tilde{A}}, \tilde{x}_{\underline{A}})\dot{\tilde{A}}^{-1} - \delta(\underline{A}, \dot{A}_{\underline{A}}, x_{\underline{A}})\dot{A}_{\underline{A}}^{-1}, \tag{71}$$

where $\tilde{x}_{\underline{A}}$ is given by (29), at $A = \underline{A}$, with $\dot{\tilde{A}}$ in place of $\dot{A}_{\underline{A}}$.

If $\underline{A} \geq A^*$, (71) reduces to

$$\big(\bar{\delta}^{1-\gamma}\beta^{\beta}\underline{A}^{(\beta-\alpha)(\gamma-1)}v_{\underline{A}}^{\beta}\big)^{-\frac{1}{\beta+\gamma-1}}\Big(\dot{\tilde{A}}^{\zeta\frac{\gamma-1}{\beta+\gamma-1}-1} - \dot{A}_{\underline{A}}^{\zeta\frac{\gamma-1}{\beta+\gamma-1}-1}\Big).$$

Since $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$, this is negative if $\zeta < 1 + \frac{\beta}{\gamma-1}$, zero if $\zeta = 1 + \frac{\beta}{\gamma-1}$, and positive if $\zeta > 1 + \frac{\beta}{\gamma-1}$.

If $\underline{A} < A^*$, so that $x_{\underline{A}} = 1$, and $\dot{\tilde{A}}$ is small enough to maintain $\tilde{x}_{\underline{A}} = 1$, then (71) reduces to

$$\bar{\delta}\underline{A}^{\alpha}\big(\dot{\tilde{A}}^{\zeta-1} - \dot{A}_{\underline{A}}^{\zeta-1}\big).$$

Since $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$, this is negative if $\zeta < 1$, zero if $\zeta = 1$, and positive if $\zeta > 1$.

B.4 Proof of Proposition 4

Suppose that $R^* \leq 1$, and, by contradiction, that we do not have $C^* = \infty$.

By the failure of $C^* = \infty$, there is an increasing and unbounded sequence of times, $t_n \to \infty$, such that $C_{t_n} \leq \overline{C}\ \forall n \geq 1$.

Consider the sequence of consumption levels $n\overline{C}\ \forall n \geq 1$. Since $n\overline{C} \to \infty$, by $R^* \leq 1$ we have

$$\lim_{n\to\infty} R(n\overline{C}) = \lim_{n\to\infty}\lim_{A\to\infty}\frac{\partial\delta}{\partial x}\big(A, \frac{n\overline{C}}{A}\big)\frac{(n\overline{C})^{\gamma}}{A\rho(\gamma-1)} \leq 1. \tag{72}$$

By D5, $\frac{\partial \delta}{\partial x}(A,x)$ weakly increases in x for any A. So

$$R(C_{t_n}) \leq R(n\overline{C})\Big(\frac{C_{t_n}}{n\overline{C}}\Big)^{\gamma} \leq R(n\overline{C})\, n^{-\gamma} \;\; \forall n, \tag{73}$$

where the first inequality follows from the fact that $n\overline{C} \geq C_{t_n}$ for each n, and the second follows from $\overline{C} \geq C_{t_n}$ for each n. By (72), $R(n\overline{C})n^{-\gamma} < 1$ for sufficiently large n, so by (73) and A4, there exists an $\underline{n}$ such that

$$\frac{\partial \delta}{\partial x}\big(A_{t_n}, \frac{C_{t_n}}{A_{t_n}}\big)\frac{C_{t_n}^{\gamma}}{A_{t_n}\rho(\gamma-1)} < 1 \;\; \forall n > \underline{n}.$$

Since v_t cannot exceed $\frac{1}{\rho(\gamma-1)}$,

$$\frac{\partial \delta}{\partial x}\big(A_{t_n}, \frac{C_{t_n}}{A_{t_n}}\big)v_{t_n} < A_{t_n}C_{t_n}^{-\gamma} \;\; \forall n > \underline{n}.$$

This is compatible with optimality only if $x_{t_n} = 1$. But this is impossible for sufficiently large n, since $C_{t_n} = A_{t_n}x_{t_n} \leq \overline{C}$ and $\lim_{n\to\infty} A_{t_n} = \infty$.

Suppose that $R^* > 1$ and, by contradiction, that $C^* = \infty$. Then there is some $\underline{C}$ such that $R(\underline{C}) > 1$:

$$\lim_{A\to\infty} \frac{\partial \delta}{\partial x}\big(A, \frac{\underline{C}}{A}\big)\frac{\underline{C}^{\gamma}}{A\rho(\gamma-1)} > 1.$$

So there is an $\underline{A}$ such that

$$\frac{\partial \delta}{\partial x}\big(A, \frac{\underline{C}}{A}\big)\frac{1}{\rho(\gamma-1)} > A\underline{C}^{-\gamma} \tag{74}$$

for all $A \geq \underline{A}$. Furthermore, because the left-hand side weakly increases in $\underline{C}$ by D5

and the right-hand side strictly decreases in $\underline{C}$, (74) holds for all $A \geq \underline{A}$ and $C \geq \underline{C}$. By A4, and the supposition that $C^* = \infty$, there is a $\underline{t}$ such that

$$\frac{\partial \delta}{\partial x}\big(A_t, \frac{C_t}{A_t}\big)\frac{1}{\rho(\gamma-1)} > A_t C_t^{-\gamma} \;\; \forall t \geq \underline{t}. \tag{75}$$

Finally, optimality requires

$$\begin{aligned} A_t^{1-\gamma} x_t^{-\gamma} &\geq \frac{\partial \delta}{\partial x_t}\big(A_t, x_t\big) v_t \;\; \forall t \\ \implies (A_t x_t)^{1-\gamma}/v_t &\geq \frac{\partial \delta}{\partial x_t}\big(A_t, x_t\big) x_t \geq \delta(A_t, x_t), \end{aligned}$$

with the final inequality holding because, by D5, $\frac{\partial \delta}{\partial x} x \geq \delta$. Given $C^* = \infty$, since v_t is upper-bounded, it follows that $\delta_t \to 0$. With $\delta_t \to 0$ and $C_t \to \infty$, v_t approaches its upper bound of $\frac{1}{\rho(\gamma-1)}$.

It therefore follows from (75) that, for sufficiently large t,

$$\frac{\partial \delta}{\partial x}\big(A_t, \frac{C_t}{A_t}\big) v_t > A_t C_t^{-\gamma}.$$

This is incompatible with optimality. Thus, if $R^* > 1$, it is impossible that $C^* = \infty$.

B.5 Proof of Proposition 5a

B.5.1 Preliminaries

It is optimal to set $x_t = 1$ as long as, at $x = 1$, the marginal flow disutility of decreasing x weakly exceeds the marginal expected utility of doing so via decreasing

the hazard rate:

$$A_t^{1-\gamma} \geq \frac{\partial \delta}{\partial x}(A_t, 1)\, v_t. \tag{76}$$

It is optimal to set $x_t < 1$ as long as (76) fails, maintaining

$$A_t^{1-\gamma} x_t^{-\gamma} = \frac{\partial \delta}{\partial x}(A_t, x_t)\, v_t \tag{77}$$

$$\implies x_t = A_t^{\frac{1-\gamma}{\gamma}} \Big(\frac{\partial \delta}{\partial x}(A_t, x_t)\, v_t \Big)^{-\frac{1}{\gamma}}. \tag{78}$$

The uniqueness of the optimal path is shown in Appendix B.1.

B.5.2 Proof that $\lim_{t \to -\infty} x_t = 1$

We will show that there exists a time $\underline{t}$ such that $v_{\underline{t}} \leq 0$. It then follows immediately that $x_t = 1$ for $t \leq \underline{t}$.

Let

$$T \equiv A^{-1}\big((\gamma - 1)^{\frac{1}{1-\gamma}}\big)$$

denote the time at which $A_T = (\gamma - 1)^{\frac{1}{1-\gamma}}$, and at which therefore $u(A_T) = -1$. If $v_T \leq 0$, the result follows immediately. Let us therefore assume that $v_T > 0$.

For $t < T$,

$$\begin{aligned} v_t &= \int_t^\infty e^{-\rho(s-t) - \int_t^s \delta_q dq} u(C_s) ds \\ &= \int_t^T e^{-\rho(s-t) - \int_t^s \delta_q dq} u(C_s) ds \;+\; e^{-\rho(T-t) - \int_t^T \delta_q dq} v_T. \end{aligned} \tag{79}$$

Since $u(C_s) \leq u(A_s) \leq -1$ for $s \leq T$, the first term of (79) is negative—indeed,

an integral over s of values which are negative for all s. The integral is shrunk in magnitude when, for all s, $u(C_s)$ is replaced with -1 and the discount factor $e^{-\rho(s-t)-\int_t^s \delta_q dq}$ replaced with its minimum value across the range, namely the discount factor at T. So

$$v_t < (t - T + v_T)e^{-\rho(T-t)-\int_t^T \delta_q dq}$$
$$\implies v_{T-v_T} < 0.$$

This proof admittedly "takes the model too literally", in assuming that technology growth has always been exponential and that therefore life was not worth living before some point in the past. Still, the dynamic it bluntly illustrates should not be controversial. When $\gamma > 1$, proportional sacrifices in consumption—decreases to x—carry greater utility costs the lower the baseline consumption level is. Early in time, the discounted value of civilization v and the baseline consumption level A were both low, so large sacrifices for safety would not have been optimal.

B.5.3 Proof that $\lim_{t\to\infty} x_t = 0$ if η_A is bounded above $1-\gamma$

Generalizing (78), whether or not the $x_t \leq 1$ constraint binds we have

$$x_t \leq A_t^{\frac{1-\gamma}{\gamma}} \left(\frac{\partial \delta}{\partial x}(A_t, x_t)\, v_t\right)^{-\frac{1}{\gamma}}. \tag{80}$$

We will show that if $\eta_A(\cdot)$ is bounded above $1-\gamma$, the right-hand side has an upper bound which falls to 0 as (by A4) $A_t \to \infty$.

Because by D1 δ is positive, by D2 and D5 we have $\frac{\partial \delta}{\partial x}(A_t, x_t) \geq \delta(A_t, x_t)$. The

right-hand side is thus bounded above by

$$A_t^{\frac{1-\gamma}{\gamma}} \big(\delta(A_t, x_t) v_t\big)^{-\frac{1}{\gamma}}. \tag{81}$$

Fixing x and v, the elasticity of this upper bound with respect to A is $(1 - \gamma - \eta_A(A, x))/\gamma$. Since this is here bounded below 0, (81) tends to 0 as $A \to \infty$. Finally, v_t is positive for all $t \geq 0$, because by A1 and A2 $A_t > 1$ for all $t \geq 0$ (rendering $v_t > 0$ feasible with $x = 1$ permanently), and v_t does not fall because sufficient precautions on new technology—e.g. banning its use—allow the consumption path to be maintained without increasing risk, by D4. Therefore, if $\eta_A(\cdot)$ is bounded above $1 - \gamma$, maintaining optimality condition (80) as $A_t \to \infty$ requires $x_t \to 0$.

B.6 Proof of Proposition 6

If $\lim_{k \to 1^+} \tilde{R}(k) < 1$, there is a $\overline{k} > 1$ such that

$$\lim_{t \to \infty} \frac{\partial \delta}{\partial x}\Big(A_t, \frac{t^{\frac{\overline{k}}{\gamma - 1}}}{A_t}\Big) \frac{t^{\frac{\overline{k}\gamma}{\gamma - 1}}}{A_t \rho(\gamma - 1)} < 1. \tag{82}$$

Choose $k \in (1, \overline{k})$. Suppose that $\nexists \underline{t} \,:\, C_t > t^{\frac{k}{\gamma - 1}} \;\; \forall t > \underline{t}$. Then there is an increasing and unbounded sequence of times, $\{t_n\} \to \infty$, such that

$$C_{t_n} \leq t_n^{\frac{k}{\gamma - 1}} \;\; \forall n \geq 1. \tag{83}$$

Observe that

$$\begin{aligned}&\lim_{n\to\infty}\frac{\partial\delta}{\partial x}\Big(A_{t_n},\frac{t_n^{\frac{k}{\gamma-1}}}{A_{t_n}}\Big)\frac{t_n^{\frac{k\gamma}{\gamma-1}}}{A_{t_n}\rho(\gamma-1)}\\ \leq&\lim_{t\to\infty}\frac{\partial\delta}{\partial x}\Big(A_t,\frac{t^{\frac{\overline{k}}{\gamma-1}}}{A_t}\Big)\frac{t^{\frac{\overline{k}\gamma}{\gamma-1}}}{A_t\rho(\gamma-1)}\cdot t^{-\frac{\overline{k}-k}{\gamma-1}\gamma}=0,\end{aligned}\tag{84}$$

where the inequality follows from the fact that, by D5, $\frac{\partial\delta}{\partial x}(A,x)$ weakly increases in x, and the limit before the $t^{-\frac{\overline{k}-k}{\gamma-1}\gamma}$ term is less than 1 by (82).

By (83), (84), and the fact that $v_t<\frac{1}{\rho(\gamma-1)}$ for all t, there is an $\underline{n}$ such that, for all $n\geq\underline{n}$,

$$\frac{\partial\delta}{\partial x}\Big(A_{t_n},\frac{C_{t_n}}{A_{t_n}}\Big)v_{t_n}<A_{t_n}C_{t_n}^{-\gamma}.$$

This is compatible with optimality only if $x_{t_n}=A_{t_n}x_{t_n}=1$. But this is impossible for sufficiently large n, by (39) and (83).

So for some $k>1$,

$$\exists\underline{t}\ :\ C_t>t^{\frac{k}{\gamma-1}}\ \ \forall t>\underline{t}.\tag{85}$$

So (85) holds for $k=1$ as well.

Given (85) for some $k > 1$, we have, for some $\underline{t}$ and some $\underline{k} \in (1, k)$, that for all $t > \underline{t}$

$$\begin{aligned} (A_t x_t)^{1-\gamma} &< t^{-k} \\ \implies \frac{\partial \delta}{\partial x}(A_t, x_t) x_t \, v_t &< t^{-k} \\ \implies \delta_t v_t &< t^{-k} \\ \implies \delta_t &< t^{-\underline{k}}. \end{aligned} \tag{86}$$

The first implication follows from the fact that $A_t^{1-\gamma} x_t^{-\gamma} \geq \frac{\partial \delta}{\partial x}(A_t, x_t) v_t$ whether or not x is interior. The second follows from the fact that $\delta < \frac{\partial \delta}{\partial x} x$ by D1 and D5. The third follows from the fact that v_t is eventually positive and does not fall to zero.

δ_t is uniformly bounded from 0 to $\underline{t}$ by $\max_{A \in [A_0, A_{\underline{t}}]} \delta(A, 1)$, which exists and is finite by the continuity of $\delta(\cdot)$ (D3). It follows from this and from (86) that $S_\infty > 0$.

If $\lim_{k \to 1^-} \tilde{R}(k) > 1$, there is a $k < 1$ and an $\underline{s}$ such that

$$\frac{\partial \delta}{\partial x}\Big(A_t, \frac{t^{\frac{k}{\gamma-1}}}{A_t}\Big) \frac{t^{\frac{k\gamma}{\gamma-1}}}{A_t \rho(\gamma - 1)} > 1 \;\; \forall t > \underline{s}. \tag{87}$$

Suppose by contradiction that $\nexists \underline{t} \;:\; C_t < t^{\frac{1}{\gamma-1}} \;\; \forall t > \underline{t}$. Then there is an increasing and unbounded sequence of times, $\{t_n\} \to \infty$, such that

$$C_{t_n} \geq t_n^{\frac{1}{\gamma-1}} \;\; \forall n \geq 1. \tag{88}$$

Observe that

$$\begin{aligned}&\lim_{n\to\infty}\frac{\partial\delta}{\partial x}\Big(A_{t_n},\frac{t_n^{\frac{1}{\gamma-1}}}{A_{t_n}}\Big)\frac{t_n^{\frac{\gamma}{\gamma-1}}}{A_{t_n}\rho(\gamma-1)}\\ \geq&\lim_{t\to\infty}\frac{\partial\delta}{\partial x}\Big(A_t,\frac{t^{\frac{k}{\gamma-1}}}{A_t}\Big)\frac{t^{\frac{k\gamma}{\gamma-1}}}{A_t\rho(\gamma-1)}\cdot t^{\frac{1-k}{\gamma-1}\gamma}=\infty,\end{aligned}\tag{89}$$

where the inequality follows from the fact that, by D5, $\frac{\partial\delta}{\partial x}(A,x)$ weakly increases in x, and the limit before the $t^{\frac{1-k}{\gamma-1}\gamma}$ term is greater than 1 by (87).

By (88), (89), and the fact that $v_t \not\to 0$, there is an n such that

$$\frac{\partial\delta}{\partial x}\Big(A_{t_n},\frac{C_{t_n}}{A_{t_n}}\Big)v_{t_n} > A_{t_n}C_{t_n}^{-\gamma}.$$

This is incompatible with optimality. So

$$\exists\underline{t}: C_t < t^{\frac{1}{\gamma-1}}\ \ \forall t>\underline{t}.\tag{90}$$

By (90) and (39), $x_t \to 0$. So there exists a $\bar{t}\geq\underline{t}$ such that, for all $t>\bar{t}$, the choice of x is interior

$$\frac{\partial\delta}{\partial x}(A_t,x_t)v_t = A_t^{1-\gamma}x_t^{-\gamma}$$

and so, by (90),

$$\frac{\partial\delta}{\partial x}(A_t,x_t)x_t\,v_t = C_t^{1-\gamma} > 1/t.$$

Since $\eta_x \equiv \frac{\partial \delta}{\partial x}\frac{x}{\delta}$,

$$\eta_x(A_t, x_t)\delta(A_t, x_t)\, v_t > 1/t \;\; \forall t \geq \bar{t}.$$

Recall that an interior choice of x_t implies that $v_t > 0$, that v is upper-bounded by $\frac{1}{\rho(\gamma-1)}$, and that $\delta_t > 0$ by D1. So $\eta_x > 0 \; \forall t \geq \underline{t}$. So if η_x is upper-bounded by $\overline{\eta_x}$,

$$\delta(A_t, x_t) > \frac{\rho(\gamma-1)}{\overline{\eta_x}} \cdot \frac{1}{t} \;\; \forall t \geq \bar{t}.$$

So $S_\infty = 0$.

B.7 Proof of Proposition 7

Choose an admissible technology path $A_{(\cdot)}$ and hazard function $\delta(\cdot)$.

B.7.1 Proof of part a

Choose $\underline{A}$, $\dot{\tilde{A}}$ with $\dot{\tilde{A}} > \dot{A}_{\underline{A}}$. Define $\tilde{x}_A[\epsilon]$ as $\tilde{x}_A$ given acceleration $\tilde{A}_{(\cdot)}[\epsilon]$, etc.

v_t is weakly increasing and continuous (indeed differentiable; see Appendix B.1) in t. Since A_t is continuous, increasing, and invertible in t, v_A is continuous and weakly increasing in A. $v_{\underline{A}+\epsilon}$ is therefore continuous and weakly increasing in ϵ.

From technology level $\underline{A}+\epsilon$ onward, the technology paths, and thus the paths of both consumption and the hazard rate, are identical under $A_{(\cdot)}$ and $\tilde{A}_{(\cdot)}$. So for any ϵ (including 0), $\tilde{v}_{\underline{A}+\epsilon}[\epsilon] = v_{\underline{A}+\epsilon}$. From this, the fact that $\tilde{v}_A[\epsilon]$ is weakly increasing in A, and the fact that $\tilde{v}_{\underline{A}}[\epsilon] \geq v_{\underline{A}}$ for all ϵ, we have that for all ϵ

$$\tilde{v}_A[\epsilon] \in [\tilde{v}_{\underline{A}}, \tilde{v}_{\underline{A}+\epsilon}] \subseteq [v_{\underline{A}}, v_{\underline{A}+\epsilon}] \quad \forall A \in [\underline{A}, \underline{A}+\epsilon]. \tag{91}$$

Then by the continuity of $v_{\underline{A}+\epsilon}$ in ϵ, for any ϵ_1 there is an $\overline{\epsilon}$ such that $|v_{\underline{A}+\epsilon} - v_{\underline{A}}| < \epsilon_1 \ \ \forall \epsilon < \overline{\epsilon}$.

Adapting (58),

$$\tilde{x}_A[\epsilon] = \min\Big(1,\, x : \frac{\partial \delta}{\partial x}(A,x)A^{\gamma-1}x^{\gamma} = \frac{1}{\tilde{v}_A[\epsilon]}\Big). \tag{92}$$

By (92) and A2, $\tilde{v}_A[\epsilon] \geq v_{\underline{A}} > 0$ for all $\epsilon \geq 0, A \in [\underline{A}, \underline{A}+\epsilon]$. By D3, the implicit function theorem, and the continuity of $\min(\cdot)$, $\tilde{x}_A[\epsilon]$ is continuous in $\tilde{v}_A[\epsilon]$. So by (91) and the sentence following it, for any ϵ_2 there is an $\overline{\epsilon}$ such that, for all $\epsilon < \overline{\epsilon}$,

$$\Big|\tilde{x}_A[\epsilon] - \min\Big(1,\, x : \frac{\partial \delta}{\partial x}(A,x)A^{\gamma-1}x^{\gamma} = \frac{1}{v_{\underline{A}}}\Big)\Big| < \epsilon_2 \quad \forall A \in [\underline{A}, \underline{A}+\epsilon].$$

Again by D3, the implicit function theorem, and the continuity of $\min(\cdot)$, the second term in the absolute value is continuous in A. So for any ϵ_3 there is an $\overline{\epsilon}$ such that, for all $\epsilon < \overline{\epsilon}$,

$$\Big|\tilde{x}_A[\epsilon] - \min\Big(1,\, x : \frac{\partial \delta}{\partial x}(\underline{A},x)\underline{A}^{\gamma-1}x^{\gamma} = \frac{1}{v_{\underline{A}}}\Big)\Big| = \big|\tilde{x}_A[\epsilon] - x_{\underline{A}}\big| < \epsilon_3 \quad \forall A \in [\underline{A}, \underline{A}+\epsilon].$$

With this uniform convergence, since

$$\tilde{X}[\epsilon] - X = \int_{\underline{A}}^{\underline{A}+\epsilon} \delta\big(A, \tilde{x}_A[\epsilon]\big)\dot{\tilde{A}}^{-1}dA - \int_{\underline{A}}^{\underline{A}+\epsilon} \delta(A, x_A)\dot{A}_A^{-1}dA,$$

since $\delta(\cdot)$ is continuous in both arguments, since x_A is continuous in A, and since $\dot{A}_A^{-1}$ is right-continuous in time and thus (by continuity and monotonicity of $A_{(\cdot)}$) in

A,

$$\begin{aligned}\Delta_{\underline{A},\dot{\tilde{A}}} &\equiv \lim_{\epsilon\to 0}\frac{\tilde{X}[\epsilon]-X}{\epsilon} = \delta(\underline{A},x_{\underline{A}})\dot{\tilde{A}}^{-1} - \delta(\underline{A},x_{\underline{A}})\dot{A}_{\underline{A}}^{-1}\\ &= \delta_{\underline{A}}\big(\dot{\tilde{A}}^{-1} - \dot{A}_{\underline{A}}^{-1}\big).\end{aligned}$$

This proves (a).

B.7.2 Proof of part b

Let $\tilde{A}_{(\cdot)}$ be an acceleration to $A_{(\cdot)}$ from $\underline{A}$ to $\overline{A}$. By the definition of an acceleration and the definition of cumulative risk,

$$\tilde{X} = X + \int_{\underline{A}}^{\overline{A}} \Big(\delta\big(A,\tilde{x}_A\big)\dot{\tilde{A}}_A^{-1} - \delta(A,x_A)\dot{A}_A^{-1}\Big)dA. \tag{93}$$

For all $A \in [\underline{A},\overline{A})$, we have $\tilde{v}_A \geq v_A$, and thus, by (92) (dropping the "$[\epsilon]$" arguments) and D5, $\tilde{x}_A \leq x_A$. D1, D2, and D5 imply that $\delta(\cdot)$ weakly increases in x, so $\delta(A,\tilde{x}_A) \leq \delta(A,x_A)$. So

$$\delta\big(A,\tilde{x}_A\big)\dot{\tilde{A}}_A^{-1} - \delta(A,x_A)\dot{A}_A^{-1} \leq \Delta_{A,\dot{\tilde{A}}_A} \quad \forall A \in [\underline{A},\overline{A}].$$

This proves (b).

B.7.3 Proof of part c

If $\overline{A} < \infty$, the integral of (93) finite. So given a technology path $A_{(\cdot)}$ for which $X = \infty$ and an acceleration to $\overline{A} < \infty$, $\tilde{X} = \infty$. This proves the first part of (c).

To prove the second part of (c), it will suffice to find a hazard function $\delta(\cdot)$ and

technology path $A_{(\cdot)}$ for which $X = \infty$ and a pair of accelerations $\tilde{A}_{(\cdot)}$ to $\overline{A} = \infty$, for one of which $\tilde{X}$ is finite and for the other of which $\tilde{X}$ is infinite. We have already encountered both.

For a case of the former, consider the hazard function $\delta(A_t, x_t) = A_t x_t$, discussed following Proposition 6. As discussed there, cumulative risk given optimal policy is then infinite for any technology path eventually bounded above zero.

For a case of the latter, consider hazard function (5)—$\delta(A_t, x_t) = \bar{\delta} A_t^{\alpha} x_t^{\beta}$—with baseline technology path $A_t = (t-1)^k$ $(t \geq 0)$ and acceleration $\tilde{A}_t = (t-1)^{\tilde{k}}$ $(t \geq 0)$, where

$$k \leq \frac{\beta + \gamma - 1}{(\alpha - \beta)(\gamma - 1)} < \tilde{k}.$$

To verify that this is an acceleration, $A_t = (t-1)^k \implies t = 1 + A_t^{\frac{1}{k}}$, so $\dot{A}_t = k(t-1)^{k-1} \implies \dot{A}_A = k A^{\frac{k-1}{k}}$, which increases in k given $A > 1$ (so, for $t > 0$).

As shown in (26), here $X = \infty$ and $\tilde{X} < \infty$.

B.8 Safety in redundancy

B.8.1 From discrete to continuous

Suppose a unit of production carries a constant flow probability $\bar{\delta}$ of triggering an existential catastrophe, so that, in the absence of any safeguards, the probability that it does not trigger a catastrophe after s units of time is $e^{-\bar{\delta}s}$. To be consistent with the discrete-time specification that the probability that it triggers a catastrophe after 1 unit of time equals p, we have $1 - e^{-\bar{\delta}} = p$ and thus $\bar{\delta} = -\log(1-p)$.

With $\frac{1-x_t}{x_t}$ units of safeguards maintained around t, since each unit multiplies the probability of a catastrophic failure per unit time by a factor $\tilde{b} \in (0, 1)$, we have that

the probability that a catastrophe is avoided until $t+s$ equals $e^{-\bar{\delta}\tilde{b}^{\frac{1-x}{x}}s}$.

The probability that $A_t x_t$ equally-safeguarded units of production all avoid catastrophe until $t+s$ is thus

$$\left(e^{-\bar{\delta}b^{\frac{1-x_t}{x_t}}s}\right)^{A_t x_t} = e^{-\bar{\delta}b^{\frac{1-x_t}{x_t}}A_t x_t\, s}. \tag{94}$$

So the probability of a catastrophe by s given locally constant A, x equals 1-(94), and the hazard rate—the probability of catastrophe per unit time—at time t precisely is

$$\delta_t \equiv \lim_{s\to 0}\left(1 - e^{-\bar{\delta}b^{\frac{1-x_t}{x_t}}A_t x_t\, s}\right)/s = \bar{\delta}A_t x_t \tilde{b}^{\frac{1-x_t}{x_t}}.$$

Letting $b \equiv -\log(\tilde{b}) > 0$ yields

$$\delta_t = \bar{\delta}A_t x_t e^{-b\frac{1-x_t}{x_t}}.$$

B.8.2 Proof of Proposition 8

By Appendix B.1, there is a unique optimal path. By the reasoning following (9), the optimal choice of x is 1 until the (unique) time at which

$$\frac{\partial u}{\partial x_t}(A_t, x_t) = \frac{\partial \delta}{\partial x_t}(A_t, x_t)\, v_t \tag{95}$$

at $x_t = 1$, after which the optimal choice of x_t is interior and maintains equality (95).

Differentiating the utility function and hazard function (44), we have

$$\begin{aligned} A_t^{1-\gamma}x_t^{-\gamma} &= \bar{\delta}A_t e^{-b\frac{1-x_t}{x_t}}\left(1+\frac{b}{x_t}\right)v_t \\ \implies \frac{1}{v_t} &= \bar{\delta}A_t^{\gamma}e^{-b\frac{1-x_t}{x_t}}\left(x_t^{\gamma} + bx_t^{\gamma-1}\right). \end{aligned} \tag{96}$$

Because v_t increases monotonically and is upper-bounded, it is asymptotically positive and constant, by the monotone convergence theorem.

We must have $C_t \to \infty$. If we do not, then there is a unbounded sequence of times t_n and a consumption level $\overline{C}$ such that

$$x_{t_n} \leq \overline{C}/A_{t_n} \quad \forall n. \tag{97}$$

Substituting (97) into (96), and recalling that $A_{t_n} \to \infty$, this would imply that the right-hand side of (96) tends to 0 across $\{t_n\}$, and thus that it is not asymptotically positive.

From (96),

$$\frac{1}{v_t} = \delta_t C_t^{\gamma-1}(1 + b/x_t).$$

Since $C_t^{\gamma-1} \to \infty$, x_t cannot be negative, and $1/v_t \not\to \infty$, it follows that $\delta_t \to 0$.

Since $C_t \to \infty$ and $\delta_t \to 0$, $v_t \to \bar{v}$.

Divide both sides of (96) by $\bar{\delta}A_0^\gamma$, and take the log and then the limit. With

$$\kappa \equiv \log\left(A_0^{-\gamma}\frac{1}{\rho(\gamma-1)\bar{\delta}}\right),$$

we have

$$\lim_{t\to\infty}\left[g\gamma t - b\frac{1-x_t}{x_t} + \log\left(x_t^\gamma + bx_t^{\gamma-1}\right)\right] = \kappa$$
$$\implies \lim_{t\to\infty}\frac{x_t}{1-x_t}t = \lim_{t\to\infty}\frac{b}{g\gamma - \kappa/t + \log\left(x_t^\gamma + bx_t^{\gamma-1}\right)/t}.$$

Other than $g\gamma$, the terms in the denominator on the right-hand side must converge

to 0. This would be avoided only if there were an unbounded sequence of times t_n across which x_{t_n} grew at least exponentially with time, which is impossible, or shrank at least exponentially with time, which would send the right-hand side of (96) to zero. So

$$\begin{aligned}
&\lim_{t\to\infty} \frac{x_t}{1-x_t} t = \frac{b}{g\gamma} \\
&\implies \lim_{t\to\infty} x_t t = \lim_{t\to\infty} (1-x_t)\frac{b}{g\gamma} = \frac{b}{g\gamma} \\
&\implies \lim_{t\to\infty} x_t \frac{g\gamma}{b} t = 1,
\end{aligned}$$

since $x_t \to 0$. It then follows from the hazard function that, in the limit, δ falls to 0 at exponential rate $-g(\gamma-1) < 0$.

C Transition dynamics for simulations

For simulating the transition dynamics, it is helpful to find $\dot{x}_t$ and $\dot{\delta}_t$ as functions of t and x_t in the regime where x is interior.

Hazard function (5), used throughout Sections 3.2–3.4 and used to simulate Figures 1 and 2, is the special case of hazard function (42), used to simulate Figure 3, with $\epsilon = 1$. The calculations below therefore apply to all simulations.

FOC:

$$\begin{aligned}
&\frac{\partial u}{\partial x_t}(A_t, x_t) = \frac{\partial \delta}{\partial x_t}(A_t, x_t) v_t \\
&\implies A_t^{1-\gamma} x_t^{-\gamma} = \bar{\delta} A_t^{\alpha} x_t^{\beta-2} \Big((\beta-1)\big(1-(1-x_t)^{\epsilon}\big) + \epsilon x_t (1-x_t)^{\epsilon-1} \Big) v_t.
\end{aligned}$$

Rearranging and differentiating gives

$$v_t = \frac{1}{\bar{\delta}} \frac{A_t^{1-\gamma-\alpha} x_t^{2-\gamma-\beta}}{(\beta-1)\big(1-(1-x_t)^{\epsilon}\big) + \epsilon x_t (1-x_t)^{\epsilon-1}} \tag{98}$$

$$\implies \dot{v}_t = v_t \Big((1-\gamma-\alpha) g + (2-\gamma-\beta)\frac{\dot{x}_t}{x_t} - \epsilon \frac{\beta - (\epsilon+\beta-1)x_t}{(\beta-1)(1-x_t)^{1-\epsilon} + 1 - \beta + (\epsilon+\beta-1)x_t} \frac{\dot{x}_t}{1-x_t} \Big). \tag{99}$$

From the first-order condition with respect to the state variable S_t,

$$\begin{aligned} \dot{v}_t &= v_t(\rho + \delta_t) - u(c_t) \\ &= v_t \Big(\rho + \bar{\delta} A_t^{\alpha} x_t^{\beta-1} \big(1 - (1-x_t)^{\epsilon}\big) \Big) - \frac{(A_t x_t)^{1-\gamma} - 1}{1-\gamma}. \end{aligned} \tag{100}$$

Substituting (98) into (99) and (100), setting the results equal, and solving for $\dot{x}_t$ yields

$$\begin{aligned} \dot{x}_t = {} & x_t \big((\beta-1)(1-x_t)^{1-\epsilon} + 1 - \beta + (\epsilon+\beta-1)x_t \big)(1-x_t) \\ & \Big((2-\gamma-\beta)\big((\beta-1)(1-x_t)^{1-\epsilon} + 1 - \beta \\ & + (\epsilon+\beta-1)x_t\big)(1-x_t) - \epsilon(\beta - (\epsilon+\beta-1)x_t)x_t \Big)^{-1} \\ & \Big(\rho + \bar{\delta} A_t^{\alpha} x_t^{\beta-1}\big(1-(1-x_t)^{\epsilon}\big) - g(1-\alpha-\gamma) - \\ & \frac{(A_t x_t)^{1-\gamma} - 1}{1-\gamma} \bar{\delta} A_t^{\alpha+\gamma-1} x_t^{\beta+\gamma-2} \big((\beta-1)(1-(1-x_t)^{\epsilon}) + \epsilon x_t (1-x_t)^{\epsilon-1} \big) \Big). \end{aligned} \tag{101}$$

Differentiating the hazard function (42) with respect to t yields

$$\dot{\delta}_t = \bar{\delta} A_t^{\alpha} x_t^{\beta} \frac{1-(1-x_t)^{\epsilon}}{x_t} \Big(\alpha g + (\beta-1)\frac{\dot{x}_t}{x_t} + \epsilon \frac{(1-x_t)^{\epsilon}}{1-(1-x_t)^{\epsilon}} \frac{\dot{x}_t}{1-x_t} \Big). \tag{102}$$

Scripts for replicating Figures 1, 2, and 3 using (101) and (102), and the estimate of

S_∞ following Figure 1, are provided here:

https://philiptrammell.com/static/ERAG_code.zip.

II Dynamic Public Good Provision under Time Preference Heterogeneity

Abstract

I explore the implications of time preference heterogeneity for the private funding of public goods. The assumption that players use a common discount rate is knife-edge: relaxing it yields substantially different equilibria, for two reasons. First, time preference heterogeneity motivates intertemporal polarization, analogous to the polarization seen in a static public good game. In the simplest settings, more patient players spend nothing early in time and less patient players spending nothing later. Second, and consequently, time preference heterogeneity gives less patient players a "first-mover advantage". Departures from the common-discounting assumption are economically significant: a patient player's payoff in equilibrium, relative to that obtained when he is constrained to act according to a higher discount rate, typically grows unboundedly as his share of the initial budget falls to zero. Finally I discuss applications of these results to the debate over legal disbursement minima.

Thanks to Ian Jewitt, Max Stinchcombe, Damon Binder, and to many researchers and visitors at GPI and Open Philanthropy for helpful corrections and comments.

1 Introduction

1.1 Motivation

The modern literature on public good games begins with Bergstrom et al. (1986). Bergstrom et al. observed that well-behaved static public good games, with each player's payoff concave in the total provision of each good, feature a unique equilibrium provision of each good. In any equilibrium, each individual is indifferent to marginal reallocations of resources among the goods she herself funds and weakly prefers reallocations to goods she funds from goods she does not. Even individuals with relatively similar preferences thus typically find themselves "polarized", in the sense that they fund entirely or almost entirely non-overlapping sets of projects. Similar observations have been made independently on other occasions, e.g. by Kalai and Kalai (2001).

An extensive literature on dynamic public good games has developed as well. A central concern of this literature is efficiency, and with minor variations, the concern is typically explored in models with a single public good, a single private good, and symmetric players. It has long been understood that in a simple continuous-time model with (near-)perfect monitoring of the individual or aggregate contribution history—or, equivalently, in a discrete-time model with lagged monitoring but (near-)fully patient players—(near-)fully efficient contribution schedules can typically obtain in subgame perfect equilibrium (McMillan, 1979; Benhabib and Radner, 1992; Battaglini et al., 2014). This is because the gains from deviation are infinitesimal relative to the losses from future punishments. At the same time, dynamic interaction can introduce channels that exacerbate crowding out relative to a static setting, in letting players signal that they will contribute little in total by having

contributed little in the past (Fershtman and Nitzan, 1991; Admati and Perry, 1991). Conversely, but by similar reasoning, if the public good exhibits increasing returns, a dynamic interaction can allow players "crowd each other in" (Marx and Matthews, 2000). In all cases, the dynamic setting interacts with strategic polarization only in the sense that it can affect how one player's expenditures on the public good crowd or fail to crowd others' out.

Relatedly, despite its size, the literature on dynamic public good games near-universally assumes that the actors under consideration act under a common discount rate. This assumption is pervasive both in the purely theoretical literature, as cited above, and in applications, such as to country-level efforts to mitigate climate change: see e.g. Dutta (2017). Fishman (2019) explores bargaining over public good provision in a dynamic setting where the players use different discount rates, and a small literature building on Sorger (2006) likewise explores the implications of bargaining under time preference heterogeneity in dynamic settings, some of which could apply to public good provision problems. Finally, the only other paper on the theory of public good games even to mention time preference heterogeneity, of which I am aware—and so the only paper to do so outside a bargaining context—is Jacobsen et al. (2017), but it is set in a static environment. Lower time preference, i.e. greater concern for the future, is simply listed as one reason why individuals may have different preferences regarding the provision of an environmental good in the present.

This paper introduces time preference heterogeneity among public good providers, and argues that a natural and important implication is "intertemporal polarization", or crowding out across time.

An analysis of dynamic public good provision under time preference heterogeneity is valuable for at least two reasons.

First, individual rates of time preference vary widely.[1] Most developed-world governments publish discounting guidelines that make explicit the discount rates they use in cost-benefit analysis for public policy, revealing unambiguously that they too act under heterogeneous rates of time preference.[2] Economists' recommendations of time preferences to use in social discounting differ substantially.[3] Philanthropists' time preferences appear to vary as well, both with each other and with those of individuals and policymakers.[4] Of course, individuals, policymakers, and philanthropists all regularly contribute to public goods to which other such parties also contribute, and these parties must all decide how to allocate their contributions over time. In doing so, they participate in dynamic public good games. Real-world dynamic public good games therefore likely exhibit substantial time preference heterogeneity.[5] Our attempts to model these games, and improve public good provision processes in light of them, will likely fail if we do not account for it.

Second, in practice, many individuals currently hold philanthropically-purposed assets in tax-exempt vehicles where they are earmarked for future charitable giving,

[1] A dated but helpful review of econometric and experimental literature on time preference heterogeneity among individuals and households can be found in Alan and Browning (2010), pp. 1252–3.

[2] Compare US Office of Management and Budget (2019) and HM Treasury (2020), for instance. This is especially relevant in the context of international contributions to global public goods, because, in the absence of strong international governance, nations must effectively engage in public good games.

[3] As surveyed by Drupp et al. (2018).

[4] Elsey and Moss (2024) survey philanthropists and small donors in the Effective Altruism community on their philanthropic preferences. Of 377 respondents who answered a question about time preference, 44% report a zero rate of pure time preference and 10% report a rate of 5% per year or more. This survey question was included at my request.

[5] Note in particular that philanthropy often involves providing consumption goods to others who also to some extent provide for themselves. Time preference heterogeneity is almost intrinsic to the interaction between these two providers. This is because beneficiaries typically face mortality risks, and temptations to impatience, that we should not generally expect to appear—and which, to some extent, empirically do not appear—in the utility function of a third-party provider. See Nesje (2024) for a model grounding this difference between households' and planners' discount rates, and Frietas-Groff and Makkar (2023) for an experimental approach on this question and review of the relevant literature.

such as donor-advised funds ("DAFs"). Assets in DAFs in particular, in the United States, currently total almost $150 billion; contributions to them have historically grown at a substantially higher rate than charitable contributions as a whole; and disbursements have not risen as quickly as contributions (National Philanthropic Trust, 2020). As we will see, this pattern can straightforwardly be explained as rational behavior by patient philanthropists given "over-spending" (from their perspective) by less patient other parties. Nevertheless, it is routinely criticized as an unjustifiable withholding of charitable funds, or even as a form of tax evasion. These criticisms recently reached new prominence in the United States with the June 2021 introduction of the Accelerating Charitable Efforts ("ACE") Act by Senators Angus King and Charles Grassley, which would impose disbursement requirements on DAFs, effectively requiring their contributors to act less patiently. Due to a lack of literature on dynamic public good provision under time preference heterogeneity, the implications of such a requirement, and of similar proposals to introduce or raise charitable disbursement minima in the United States and elsewhere, have not undergone thorough economic scrutiny.

I consider a simple model in which there is a single public good, and each player's flow utility depends only on total flow spending on the good. Two players with constant but different rates of time preference make decisions about how much to contribute to the good, and how much to invest for future contribution, over an infinite horizon in continuous time.

To isolate the implications of time preference heterogeneity, I assume that the present value of each player's total contribution is fixed. Contributors decide only the schedule on which to deploy their spending, not the extent to which they will spend on public as opposed to private goods each period. As a result, the model

does not resemble the existing literature on dynamic public good provision so much as Bergstrom et al.'s original paper. The relevant change is that I explore what happens when individuals choose their contribution levels for an infinite stream of public goods over which their respective preferences differ—i.e. funding at t, for all $t \geq 0$—*in sequence*, rather than simultaneously.

The model, despite its simplicity, allows us to draw some important and broad conclusions about the implications of heterogeneous discounting for public good provision.

First, the common discounting assumption is knife-edge: even slight time preference differences, even (indeed especially) by small players, give rise to very different equilibria. In particular, time preference differences generate uniquely simple equilibria in which spending is "polarized" in the sense above, with the impatient exclusively responsible for public good funding before some date the patient exclusively responsible after. Furthermore, such equilibria are "asymmetric": they give impatient players a Stackelberg-like first-mover advantage with no analogue in a common-discounting setting.

Second, time preference heterogeneity is highly payoff-relevant. The equilibria that obtain under heterogeneous discounting can offer payoffs, at least for unusually patient parties, which differ dramatically from the payoffs they achieve when constrained to spend as would seem optimal in the absence of time preference heterogeneity.

1.2 Related discounting literature

Though time preference heterogeneity has not been studied in the setting of dynamic public good provision by private funders, three strands of literature have studied time

preference heterogeneity in adjacent settings. The results of the present paper prove complementary to the central findings in each of these settings.

One strand concerns the collective allocation of private consumption over time—i.e., under certain preference aggregability assumptions, the discounting behavior of a representative agent—in a population of households with heterogeneous time preferences. A classic observation from this literature is that consumption (Rader, 1981) and/or wealth (Becker, 1980; Ryder, 1985) can, in the limit, become entirely concentrated in the hands of society's most patient members simply because they consume less and invest more. Another observation, closely related to the first, is shown in an exchange economy by Gollier and Zeckhauser (2005) and with variations elsewhere: that given complete markets, a representative agent, if one exists, will exhibit a discount rate that declines with time to that of society's most patient members. In other words, interest rates fall as patient parties lend to their less patient counterparts and command an ever-growing share of the financial market. Heal and Millner (2014) argue that policymakers aiming to set discounting policy for the provision of a public good, while deferring to the time preferences of their constituents, do best to defer to this aggregated discounting schedule. As we will see, similar dynamics play out among agents spending on public goods directly, and the strategic logic of intertemporal polarization can make the tendency for wealth to concentrate in patient hands—and the corresponding tendency of a representative agent to grow more patient with time—even more extreme.

A second body of relevant research concerns optimal taxation by policymakers more patient than their constituents. Farhi and Werning (2007, 2010) analyze optimal taxation in an intergenerational model where individuals save insufficiently, from the patient social planner's perspective, for their descendants. Household con-

sumption in these models is a public good: its provision satisfies the preferences of multiple parties (the policymaker and the household itself) nonexcludably and nonrivally. Similarly, von Below (2012), Belfiori (2017), and Barrage (2018) study optimal carbon taxation and/or investment subsidization in contexts where present production confers both future costs and future benefits (from climate damage and capital accumulation respectively). An important lesson from this literature is that patient policymakers might like to invest resources for future spending, but that to avoid crowding out private investment, it is often optimal for them instead to subsidize private investment and tax private consumption.

Time preference heterogeneity has different implications in the context of optimal taxation than in the context of private spending on public goods, however. The former setting involves an asymmetry in the players' strategy sets: households cannot tax or subsidize policymakers, but policymakers can tax and subsidize households. At least in the absence of political or informational constraints, policymakers endorsing a given time preference rate can often use these tools to implement population-wide behavior that is optimal or near-optimal from the policymaker's perspective.

Finally, a small literature considers public good provision over time by committees of social planners with different time preference rates. One finding is that any attempt to define and implement "optimal" spending plans by such a committee faces some variety of the preference aggregation impossibilities faced in other social choice contexts, as explored in detail by Chambers and Echenique (2018). Millner (2020) proposes a method by which the discounting planners might reach a kind of consensus, but such proposals are themselves inevitably vulnerable to disagreement. Another concern of the literature is that, once an optimality criterion or consensus has been reached, the resulting spending plans are typically time inconsistent. In

particular, Jackson and Yariv (2015) show that any social welfare function used in this setting must be either dictatorial or time inconsistent, in that future committee meetings will, if they use the same forward-looking social welfare function, decide to revise the plans made by previous meetings—at least if these were made naively, without taking the possibility of future revisions into account. Millner and Heal (2018) therefore examine the collective decision-making of discounting committees aware that they are playing a dynamic game with their future selves. They find that attempts to implement weighted utilitarian social discounting in such a dynamic game will generally be inefficient. By contrast, I find that decentralized private actors strategically allocating public good contributions over time can implement efficient, weighted utilitarian social discounting.

1.3 Outline

The structure of this paper is as follows.

Section 2 opens with three benchmark settings to which the dynamic public good game is later compared. The first is that of dynamic private good provision. The second and third are static transformations of a dynamic public good game: an "open-loop" setting in which the parties simultaneously commit to a spending schedule, and a "Stackelberg" setting in which the less patient player commits to a spending schedule and the patient player then chooses his best response. The former illustrates the intertemporal polarization, and the latter illustrates both the intertemporal polarization and the first-mover advantage, observed in the dynamic game.

In Section 3, I explore the dynamic game. I find that it has many equilibria, including efficient equilibria, but only one that is polarized; and that this polarized

equilibrium is also the only equilibrium that is continuous with equilibrium behavior over a lengthening finite horizon. It implements the profile of spending schedules observed in the equilibrium of the static Stackelberg game.

In Section 4, I introduce constraints in the form of a minimum and/or maximum on the proportional rate at which a player may spend. I then illustrate the economic significance of time preference heterogeneity by calculating a player's willingness to pay to move from the maximally polarized equilibrium of the constrained game to the polarized equilibrium of the unconstrained game of Section 3. I find that, for an atypically patient player (but not for an atypically impatient player), this willingness to pay approaches the entirety of his budget as his budget share—his fraction of the sum of the parties' budgets—goes to zero. That is, when most of the funding for his chosen cause is governed impatiently, a patient player finds a disbursement minimum approximately as costly as a total expropriation.

Finally, Section 5 briefly shows that intertemporal polarization is not an artefact of the simple preference specification of the games above, in which each player's flow utility is isoelastic in total flow spending, but holds under much weaker preference conditions. Since intertemporal polarization largely drives the subsequent results (on the impatient first-mover advantage and the costs of spending constraints), this suggests that the conclusions of Section 3 and Section 4 are relatively robust.

Section 6 concludes.

2 Benchmarks

To build our intuitions about the implications of time preference heterogeneity among public good funders, we consider three benchmarks.

First, in Section 2.1, we review the implications of time preference heterogeneity

among households purchasing private goods, making the standard observation that optimal spending rates rise (i.e. optimal saving rates fall) continuously with rates of time preference.

In Section 2.2, we study a static game in which two players with different discount rates simultaneously set schedules on which they will spend on a public good. (It is the "open loop" transformation of the dynamic public good game studied in the next section.) This game has a unique and fully polarized Nash equilibrium, with only the less patient player spending early in time and only the more patient spending later in time. It illustrates how time preference differences can motivate more extreme differences in saving behavior in the context of public good funding than in the context of private good funding.

The last benchmark, in Section 2.3, is a two-period game, like the static game above but with the less patient player as Stackelberg leader: choosing her spending schedule before the more patient player chooses his. The game has a unique and fully polarized subgame perfect equilibrium, like the Nash equilibrium of the (fully) static game but with the impatient player using her first-mover advantage to shift spending toward the present. It illustrates that the strategic polarization observed in a static setting can be found in (at least something closer to) a true dynamic setting, but that the dynamics of a public good game can also introduce an asymmetry between the positions of more and less patient parties.

The dynamic game of the next section, which lies at the heart of this paper, will be used to explore this polarization and this asymmetry in more detail.

Uniqueness and measure-zero deviations

A household's optimal spending schedule, and the equilibria of the benchmark public good games, were described above as unique. Strictly speaking, since we will work in continuous time and assume that measure-zero deviations in spending rates do not affect payoffs, they are unique only up to measure-zero deviations.

Throughout the paper, we will take the standard approach of restricting our attention to right-continuous spending schedules—i.e. spending schedules that are right-continuous everywhere, and continuous everywhere except at a finite set of jump discontinuities—in both individual optimization decisions and strategic interactions. Since distinct right-continuous spending schedules differ from each other at a positive-measure set of times, a right-continuous spending schedule that satisfies some property uniquely up to measure-zero deviations also satisfies it uniquely under a right-continuity restriction. However, the "unique" optimal spending schedules and equilibria found in Propositions 1 and 3–5 are unique up to measure-zero deviations, not merely uniquely under the restriction of right-continuity. This is shown in the proofs of these propositions, in the relevant appendices.

The proofs of some later propositions use right-continuity, so the assumption of right-continuity is imposed throughout the body of the paper for consistency.

2.1 Private goods

Recall the familiar "cake-eating" problem of a household that is the sole provider of its own consumption over an infinite horizon. Denote the size of the household's budget at time $t = 0$ by B. Assume that flow utility u at time t is an isoelastic function,[6] with inverse elasticity of intertemporal substitution $\gamma > 0$, of the spending

[6] We assume isoelasticity until the generalization of Section 5 because time-separable preferences (104) are homothetic if and only if flow utility is isoelastic. Homotheticity is desirable here because

rate $x_t \geq 0$:

$$u(x_t) = \begin{cases} \frac{x_t^{1-\gamma}}{1-\gamma}, & \gamma \neq 1; \\ \ln(x_t), & \gamma = 1. \end{cases} \tag{103}$$

The household faces a constant instantaneous real interest rate r and a constant instantaneous time preference rate δ. The household's problem is then to choose a right-continuous schedule (i.e. spending schedule) $x = \{x_t\}_{t \geq 0}$ that maximizes

$$\int_0^\infty e^{-\delta t} u(x_t) dt \tag{104}$$

subject to the budget constraint

$$\int_0^\infty e^{-rt} x_t dt \leq B. \tag{105}$$

Let

$$\alpha \equiv \frac{r\gamma - r + \delta}{\gamma}. \tag{106}$$

Proposition 1. ***Optimal private schedule***

Under budget constraint (105), if $\alpha > 0$, utility function (103)–(104) is uniquely maximized by schedule

$$x_t = B\alpha e^{(r-\alpha)t}, \quad t \geq 0. \tag{107}$$

If $\alpha \leq 0$, there is no optimal schedule.

Proposition 2, and the analogous "proportional willingness to pay" results of Section 4.3 that are at the heart of this paper, require it. Though similar results may obtain under weaker conditions, the inability to discuss proportional costs and benefits would come with some loss in clarity.

Proof. Proofs may be found in many introductions to dynamic optimization (see e.g. Barro and Sala-i-Martin (2004), ch. 2.1), but for convenience, one may be found in Appendix A.1, along with an expression for the payoff to following the optimal schedule. □

Note that, since $\gamma > 0$ by assumption, the condition that $\alpha > 0$ is equivalent to the condition that

$$\delta > r(1 - \gamma). \tag{108}$$

If (108) holds, the optimal spending rate as a proportion of the budget at any time is constant and equal to α.

The optimal spending rate is sensitive to the discount rate: the lower δ is, the lower α is, and the more slowly it is optimal to spend. In fact, when (108) is violated, a lower proportional spending rate is always preferable to a higher one. Since a permanent spending rate of zero is of course worst of all, an optimal schedule does not exist. Note that (108) does hold whenever $\gamma > 1$, $r > 0$, and $\delta \geq 0$. That is, under the standard assumptions that $r > 0$ and $\gamma > 1$, an optimal schedule exists even under full patience.

As (107) reveals, the optimal private spending rate is continuous in δ, throughout the region in which an optimum exists. It is shown below that this continuity does not generally hold in the setting of public good provision.

Costs of over- or under-spending

Now imagine that a household with time preference rate δ is required to spend its budget as would be optimal according to some time preference rate $\tilde{\delta} \neq \delta$, fixing r and γ. Assume that δ and $\tilde{\delta}$ both satisfy (108), as above, but here further assume

that

$$\begin{aligned} &\gamma \leq 1 \\ \text{or} \quad &\tilde{\delta} < \delta \frac{\gamma}{\gamma - 1} + r. \end{aligned} \tag{109}$$

If neither part of this condition is satisfied, the optimal schedule given $\tilde{\delta}$ sends spending to zero quickly enough that, over the infinite horizon, it produces infinite δ-discounted disutility. Note that $\tilde{\delta} < \delta + r$ is sufficient for the condition to hold.

Let $\tilde{\alpha}$ be defined as α is in (106), but with $\tilde{\delta}$ in place of δ. Then define

$$\eta \equiv \begin{cases} \left(\frac{\tilde{\alpha}+\delta-\tilde{\delta}}{\alpha}\right)^{\frac{1}{1-\gamma}}, & \gamma \neq 1; \\ \\ e^{\frac{\tilde{\delta}}{\delta}-1}, & \gamma = 1. \end{cases} \tag{110}$$

η may be interpreted as a measure of the extent to which $\tilde{\delta}$ exceeds δ. Observe that $\eta > 1$ when $\tilde{\delta} > \delta$ and $\eta < 1$ when $\tilde{\delta} < \delta$.

Finally, let $w(B)$ denote the household's willingness to pay to avoid the spending rate requirement, as a proportion of its budget B.

Proposition 2. ***Bounded WTP for optimal private schedule***
$w(B) = 1 - \frac{\tilde{\alpha}}{\alpha\eta} < 1$, *independent of* B.

Proof. A proof may be found in Appendix A.2, along with an expression for the payoff to spending $\tilde{\delta}$-optimally. □

So the spending rate requirement—the requirement to spend at proportional rate $\tilde{\alpha}$ rather than α—lowers the household's utility by as much as a tax of proportion $1 - \frac{\tilde{\alpha}}{\alpha\eta}$ of its budget. As shown in Section 4.3, this cost can be greatly magnified in

the setting of public good provision. In particular, as a patient player grows small, his proportional willingness to pay to avoid a spending rate minimum generally rises to 1.

2.2 Open-loop game: intertemporal polarization

Consider the following static game, concerning the allocation of spending on a public good across periods.

There are two players, H and L. Each player i chooses a piecewise right-continuous individual schedule $x^i \equiv \{x_t^i\}_{t\geq 0}$ subject to the budget constraint

$$\int_0^\infty e^{-rt} x_t^i dt \leq B^i.$$

A pair of individual schedules (x^H, x^L), called simply a "schedule", is denoted x. The sum of the players' schedules $x^H + x^L$ is called a "collective schedule" and denoted X.

H and L share an isoelastic flow utility function $u(\cdot)$, but now flow utility is a function of total flow spending by both parties: the good being spent on is a public good. The only difference between the players' preferences is that they discount at different rates. H discounts at the higher rate:

$$\delta^H > \delta^L > r(1-\gamma). \tag{111}$$

Thus i's payoff, as written as a function of the schedule, is

$$U^i(x) = \int_0^\infty e^{-\delta^i t} u\big(x_t^H + x_t^L\big) dt.$$

$U^i(\cdot)$ may also be written as a function of a collective schedule.

Let U denote a payoff profile (U^H, U^L). Define α^i as in (106), with δ^i in place of δ.

This game is the open-loop transformation of the dynamic public good game studied throughout the next section, so we will call it the "open-loop game".

Proposition 3. ***Existence and uniqueness of Nash equilibrium in the open-loop game***

The game above has a unique Nash equilibrium $x^{(o)}$*:*

$$x_t^{H(o)} = \begin{cases} B^H \alpha^H \, \frac{M^o}{M^o - 1} \, e^{(r-\alpha^H)t}, & t < t^o; \\ 0, & t \geq t^o, \end{cases} \tag{112}$$

$$x_t^{L(o)} = \begin{cases} 0, & t < t^o; \\ B^L \alpha^L \, M^o \, \frac{\alpha^L}{\alpha^H} \, e^{(r-\alpha^L)t}, & t \geq t^o, \end{cases} \tag{113}$$

where

$$t^o \equiv \ln(M^o) / \alpha^H, \tag{114}$$

$$M^o \equiv 1 + \frac{B^H \alpha^H}{B^L \alpha^L}. \tag{115}$$

Proof. See Appendix A.3. □

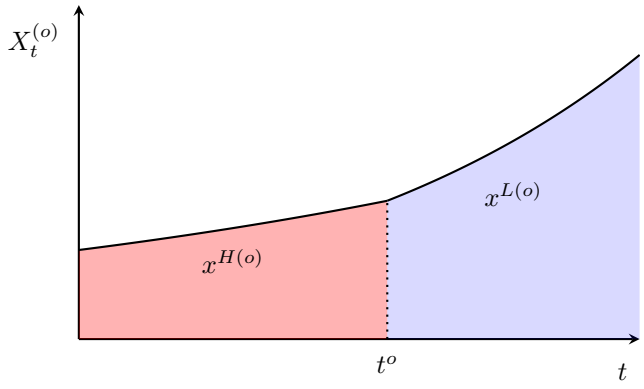


Fig. 1: Equilibrium schedule of the open-loop game

This game is essentially a special case of the static public good game analyzed by Bergstrom et al. (1986), but with a continuum of public goods: spending at each t, for $t \in [0, \infty)$. As in the analogous Bergstrom et al. case, each good here is provided by exactly one funder, and when a funder provides a good, she always also is the provider of all goods about which she cares relatively more. Since L cares relatively more than H about spending at t the later t is, there is a threshold time t^o such that H is the sole funder before t^o and L is the sole funder after. Unlike in the case of the private good, therefore, even a slight time preference difference motivates a big difference in behavior between the public good funders.

We can also see from (112) and (113) that i's spending growth rate here equals $r - \alpha^i$ across the period during which i spends. Since $r - \alpha^i = (r - \delta^i)/\gamma$, by (107) this is i's optimal spending growth rate: the growth rate i chooses when she is the only funder. If i's spending growth rate were not δ^i-optimal across t with $x_t^i > 0$, $x^{i(o)}$ would not be an equilibrium strategy; i would prefer marginal reallocations of funding across periods.

Finally, substituting (114) into (112) and (113), we see that the collective spending rate is continuous at t^o. That is, $\lim_{t \to t^{o-}} x_t^{H(o)} = x_{t^o}^{L(o)}$. If the spending rate rose discontinuously at t^o, L would do better to reallocate some spending from $t^o + \epsilon$ to

$t^o - \epsilon$ for some sufficiently small $\epsilon > 0$. Likewise, if the spending rate fell discontinuously, H would do better to reallocate marginal spending forward. The relative budget sizes, the spending growth conditions, and the continuity condition pin down t^o and thus the equilibrium.

2.3 Stackelberg game: first-mover advantage

Finally, consider a two-period game, identical to the game above except that H is the "Stackelberg leader". That is, suppose that, in the first period, H sets a feasible spending schedule x^H, and in the second period, L observes x^H and sets a feasible spending schedule x^L in response. H's strategy set is thus the set of piecewise right-continuous spending schedules satisfying the budget constraint imposed by B^H, and L's strategy set is the set of functions from feasible choices of x^H to feasible choices of x^L.

Define η as in (110), with δ^H in place of δ and δ^L in place of $\tilde{\delta}$. Since $\delta^L < \delta^H$, we have $\eta \in (0, 1)$.

Proposition 4. ***Existence and uniqueness of SPE in the Stackelberg game***

The game above has a unique subgame perfect equilibrium, which implements the schedule x^*:

$$x_t^{H*} = \begin{cases} B^H \alpha^H \frac{M^*}{M^* - 1} e^{(r-\alpha^H)t}, & t < t^*; \\ 0, & t \geq t^*, \end{cases}$$

$$x_t^{L*} = \begin{cases} 0, & t < t^*; \\ B^L \alpha^L M^{* \frac{\alpha^L}{\alpha^H}} e^{(r-\alpha^L)t}, & t \geq t^*, \end{cases}$$

where

$$t^* \equiv \ln(M^*)/\alpha^H, \tag{116}$$
$$M^* \equiv 1 + \frac{B^H \alpha^H}{B^L \alpha^L}\,\eta.$$

Proof. See Appendix A.4. □

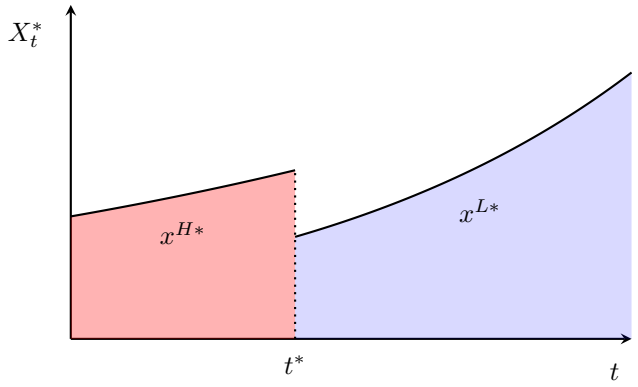


Fig. 2: Equilibrium schedule of the Stackelberg game

Since $\eta < 1$, $t^* < t^o$: the regime-switching time occurs earlier in the Stackelberg case than in the case where the players set their spending plans simultaneously. Furthermore, recall that in the open-loop (i.e. simultaneous-move) case, spending is continuous at t^o. Here, the spending rate falls discontinuously at t^*, as H allocates budget B^H over a shorter time interval and L allocates B^L over an infinite horizon beginning earlier.

$U^H(x^*) > U^H(x^{(o)})$. This follows immediately from the uniqueness of the SPE in the Stackelberg game, and from the fact that H can attain the open-loop schedule in it: if H chooses $x^H = x^{H*}$, L's best response is x^{L*} by Proposition 3. Conversely, $U^L(x^o) > U^L(x^*)$. This follows from the fact that X^* can be obtained by beginning with $X^{(o)}$, decreasing funding at times $t \geq t^*$, and increasing funding at times $t < t^*$. L disprefers any marginal reallocation of funding from times after t^* to times before

t^* from the $X^{(o)}$ baseline, and because $u(\cdot)$ is concave, L only disprefers non-marginal reallocations the more strongly. In short, the Stackelberg game gives H a first-mover advantage.

3 Dynamic game

The open-loop and Stackelberg games suggest two ways in which time preference differences can shape public good provision. The first is by motivating what might be called *intertemporal polarization*. The second is by giving impatient parties a *first-mover advantage*. We will now see how these stylized results obtain in a more realistic dynamic setting, in which players can continuously observe each other's spending rates and update their own future spending plans. Equilibrium behavior in continuous time is defined precisely in Section 3.1, and the two implications above are detailed in Section 3.2, under an equilibrium refinement.

Unlike the benchmark settings, however, the dynamic setting admits a wide range of equilibria. This is discussed in Section 3.3.

3.1 Setup

Equilibrium behavior in the continuous-time dynamic setting will be defined as a limit of equilibrium behavior along a sequence of dynamic games with ever more frequent discrete-time monitoring, indexed by $n \geq 0$. A justification for this modeling assumption (and other complications that must be introduced) is given below, under "Equilibria and motivations".

As in the benchmark games, there are two players, H and L. Player i begins with budget $B^i > 0$ and faces a constant instantaneous real interest rate r. Player

i's realized spending rate at t is denoted x_t^i, and i's payoff is

$$U^i(x) \equiv \int_0^\infty e^{-\delta^i t} u(x_t^H + x_t^L) dt, \tag{117}$$

where $u(\cdot)$ is isoelastic, as given by (103), and the time preference rates satisfy (111).

To summarize the section: game n partitions the non-negative real line into a (perhaps infinite) number of periods of positive length. At the beginning of each period, each player observes the history of play across previous periods and publicly chooses a time weakly before the end of the period at which she will announce her spending history since the start of the period. Each player then sets a spending plan until the first announcement. When i announces, $-i$ observes the history and sets a spending plan to the end of the period. If $-i$ has not yet announced in this period, then she also publicly plans an announcement at some time weakly before the end of the period, and i sets a spending plan until $-i$'s announcement.

Schedules

A schedule for i, denoted x^i, is again a spending rate for i at every time $t \geq 0$:

$$x^i \equiv \{x_t^i\}_{t \geq 0}.$$

Given an interval $I \subset [0, \infty)$ that is closed below and open above, a truncated schedule for i, denoted x_I^i, is a spending rate for i at every time $t \in I$:

$$x_I^i \equiv \{x_t^i\}_{t \in I}.$$

A truncated schedule for i with $I = [0, t)$, for finite t, is a partial schedule for i and may be denoted $x^i_{|t}$.

A schedule $x \equiv (x^H, x^L)$ is a pair of player schedules. A truncated or partial schedule, x_I or $x_{|t}$, is a pair of such player schedules over the same interval.

Given a schedule x, a collective schedule $X \equiv x^H + x^L$ is a total spending rate at every time $s \geq 0$. Truncated and partial collective schedules are defined and denoted analogously to such player schedules.

Given a truncated schedule x_I,

$$U^i(x_I) \equiv \int_I e^{-\delta^i(t-\min(I))} u\big(x^H_t + x^L_t\big) dt.$$

$U^i(\cdot)$ is also defined over collective [truncated] schedules. Payoff profiles $U(\cdot)$ are pairs of individual payoffs, as previously.

A partial schedule $x_{|t}$ is feasible if it is right-continuous and

$$\int_0^t e^{-rs} x^i_s ds \leq B^i \ \ \forall i.$$

The feasibility of a schedule is defined likewise.

Given a partial schedule $x_{|t}$,

$$B^i_{\underline{t}}(x_{|t}) \equiv \Big(B^i - \int_0^{\underline{t}} e^{-rs} x^i_s ds\Big) e^{r\underline{t}}$$

denotes i's implied budget at time $\underline{t} \leq t$. Given a schedule x, $B^i_{\underline{t}}(x)$ can be defined likewise, with no restriction on $\underline{t}$. For simplicity, $B^i(x_{|t}) \equiv B^i_t(x_{|t})$.

$B = B^H + B^L$ denotes the (initial) collective budget, and $b \equiv B^L/B$. Like the B^i, B and b can be written with time subscripts or as functions of a [partial] schedule.

Strategies

In each game, a set of times are called grid points. The set of grid points in game n is denoted $G^{(n)}$ and includes 0. The grid sequence $\mathcal{G} \equiv \{G^{(n)}\}_{n=0}^{\infty}$ is such that

- for all n, $G^{(n+1)} \supset G^{(n)}$ and $G^{(n)}$ is locally finite (so that $G^{(n)} \cap [0,T]$ is finite for all T); and
- $\cup_{n=0}^{\infty} G^{(n)}$ is dense in $\mathbb{R}_+$.

Given a time $t \geq 0$,

$$\tau^{(n)}(t) \equiv \max\left(G^{(n)} \cap [0,t]\right)$$

denotes the largest grid point in game n less than or equal to t, and $\tau^{(n)\prime}(t)$ denotes the smallest grid point in game n greater than t. Given $\tau \in G^{(n)}$, $\tau^{(n)\prime} \equiv \tau^{(n)\prime}(\tau)$ denotes the subsequent grid point in game n.

A period of game n is an interval $[\tau, \tau^{(n)\prime})$ with $\tau \in G^{(n)}$.

Henceforth, where doing so will come with no loss of clarity, we will fix n and drop the "(n)" superscripts.

An announcement ξ_τ^i is a time associated with a player i and a grid point τ with $\xi_\tau^i \in [\tau, \tau')$.

A node $h_{|t}$ is a history of play up to t: a feasible partial schedule across $[0,t)$, denoted $x(h_{|t})$, paired with a set of announcements $\{\xi_\tau^i(h_{|t})\}$ that includes exactly one entry-pair (one entry for each player) for each $\tau \in G \cap [0,t)$ and none for $\tau > t$, and such that (i) if $t \in G$ and $\xi_t^i(h_{|t})$ is defined for both i, then $\xi_t^H(h_{|t}) = \xi_t^L(h_{|t})$; and (ii) if $t \notin G$, $\xi_{\tau(t)}^i(h_{|t}) = t$ for some i.

If no announcement associated with some i, τ is contained in $\{\xi_\tau^i(h_{|t})\}$, we write $\xi_\tau^i(h_{|t}) = \emptyset$. Observe that every node $h_{|t}$ is of one of five types:

1. $t \in G$, $\xi_t^i(h_{|t}) = \emptyset$ for both i.

2. $t \in G$, $\xi_t^i(h_{|t}) > t$ for one or both i (and $= \emptyset$ otherwise).

3. $t \notin G$, $\xi_{\tau(t)}^i(h_{|t}) = \emptyset$ for one i (and $\xi_{\tau(t)}^{-i}(h_{|t}) = t$).

4. $t \notin G$, $\xi_{\tau(t)}^i(h_{|t}) > t$ for one i (and $\xi_{\tau(t)}^{-i}(h_{|t}) = t$).

5. $t \notin G$, $\xi_{\tau(t)}^i(h_{|t}) \leq t$ for both i (and $= t$ for one or both i).

The interpretation of these nodes is made clearer by the sequence of play outlined just below.

A node of type 1 or 3 is a pre-announcement node. A node of type 2, 4, or 5 is a post-announcement node. Given a post-announcement node $h_{|t}$,

$$\begin{aligned} \hat{\xi}(h_{|t}) &\equiv \max_{i,\tau}\{\xi_\tau^i(h_{|t})\} \quad \text{if } h_{|t} \text{ is of type 2 or 4,} \\ &\equiv \tau'(t) \qquad\qquad\quad \text{if } h_{|t} \text{ is of type 5} \end{aligned}$$

denotes the time of the next node. We will sometimes distinguish pre- and post-announcement nodes occurring at the same time by denoting the former $h_{|t}^-$ and the latter $h_{|t}^+$.

The game begins with a trivial node of type 1 at $t = 0$.

Individual budgets $B^i(\cdot)$, the collective budget $B(\cdot)$, or L's budget share $b(\cdot)$ may be written as functions of nodes rather than of their associated partial schedules.

The set of nodes is denoted $\mathcal{H}$. We will say for simplicity that every element of $\mathcal{H}$ is a node for each player, though as we will see, some players take no substantive action at some nodes.

A strategy for i in game n, denoted σ^i, is a function that maps $h_{|t} \in \mathcal{H}$ to one of the following actions. The list below also defines how the subsequent node follows from the action profile $\sigma(h_{|t})$.

- If $h_{|t}$ is of type 1: $\sigma^i(h_{|t})$ is an announcement in $(t, t']$.

  The subsequent node occurs at t and is of type 2, and is $h_{|t}$ appended with $\xi_t^j = \sigma^j(h_{|t})$ if $\sigma^j(h_{|t}) \leq \sigma^{-j}(h_{|t})$ for each j.

- If $h_{|t}$ is of type 2: $\sigma^i(h_{|t})$ is a feasible spending plan to the time of the next node $\hat{\xi}(h_{|t})$.

  In general a feasible spending plan from node $h_{|t}$ to $\bar{t} > t$ is a truncated schedule for i, $x^i_{[t,\bar{t})}$, that is right-continuous and satisfies the budget constraint

$$\int_t^{\bar{t}} e^{-r(s-t)} x_s^i ds \leq B^i(h_{|t}). \tag{118}$$

  If $\hat{\xi}(h_{|t}) = t'$, the subsequent node occurs at t' and is of type 1. If $\xi_t^H(h_{|t}) = \xi_t^L(h_{|t})(= \hat{\xi}(h_{|t})) < t'$, the subsequent node occurs at $\hat{\xi}(h_{|t})$ and is of type 5. Otherwise, the subsequent node occurs at $\hat{\xi}(h_{|t})$ and is of type 3. In all cases, the subsequent node is $h_{|t}$ with the associated partial schedule appended with the players' spending plans.

- If $h_{|t}$ is of type 3...

  a. and $\xi^i_{\tau(t)}(h_{|t}) = \emptyset$: $\sigma^i(h_{|t})$ is an announcement $\sigma^i(h_{|t}) \in (t, \tau'(t)]$.

  b. and $\xi^i_{\tau(t)}(h_{|t}) = t$: $\sigma^i(h_{|t}) = \emptyset$.

  Let j denote the player whose $\tau(t)$-announcement at $h_{|t}$ was undefined. The subsequent node occurs at $\sigma^j(h_{|t})$ and is of type 4, and is $h_{|t}$ appended with j's announcement.

- If $h_{|t}$ is of type 4: $\sigma^i(h_{|t})$ is a feasible spending plan to $\hat{\xi}(h_{|t})$.

  If $\hat{\xi}(h_{|t}) = \tau'(t)$, the subsequent node occurs at $\tau'(t)$ and is of type 1. Otherwise, the subsequent node occurs at $\hat{\xi}(h_{|t})$ and is of type 5. In either case, the subsequent node is $h_{|t}$ with the partial schedule appended with the players' spending plans.

- If $h_{|t}$ is of type 5: $\sigma^i(h_{|t})$ is a feasible spending plan to $\tau'(t)$.

  The subsequent node occurs at $\tau'(t)$ and is of type 1, and is $h_{|t}$ appended with the players' spending plans.

i's strategy set (in game n, though we will continue to suppress the superscript) is denoted Σ^i. A strategy profile (σ^H, σ^L) is denoted σ. The set of strategy profiles is denoted Σ. We will not introduce mixed strategies.

At $h_{|t} \in \mathcal{H}$, if players adopt strategy profile σ, the resulting schedule is denoted $x(h_{|t}, \sigma)$. For simplicity, $x(\sigma) \equiv x(\emptyset, \sigma)$ denotes the schedule generated by σ. The collective schedule resulting from σ following $h_{|t}$ is likewise denoted $X(h_{|t}, \sigma)$, etc. $h^-_{|s}(h_{|t}, \sigma)$ / $h^+_{|s}(h_{|t}, \sigma)$ is the pre-/post-announcement node implemented by σ at $s \geq t$ given $h_{|t}$, if such a node exists. (Note that it must if $s \in G$.)

Equilibria and motivations

We will take the natural equilibrium concept to be SPE. A strategy profile $\sigma^* \in \Sigma$ is an equilibrium (of game n) if

$$U^i\big(x_{[t,\infty)}(h_{|t}, \sigma^*)\big) \geq U^i\big(x_{[t,\infty)}(h_{|t}, (\sigma^i, \sigma^{*-i}))\big) \quad \forall i \; \forall h_{|t} \in \mathcal{H}. \tag{119}$$

Though including announcements makes almost no difference to the structure of the game when n is large, it simplifies our efforts to characterize its equilibria.

Without announcements, in an equilibrium of game n in which H exhausts her budget, she typically does so at a grid point—an element of $G^{(n)}$—because she places some value on L immediately observing her budget-exhaustion and beginning to spend. So as we vary parameters of the game (e.g. as we vary the initial budgets at the node initiating a subgame, by exploring previous-period deviations), the time at which H chooses to exhaust her budget typically jumps between elements of $G^{(n)}$. Allowing the players to make announcements removes the additional value to H of spending down on a grid point, letting her equilibrium spend-down point vary continuously with the history (and the parameter values) of the game.

The complications induced by H's desire to spend down at a grid point could be addressed in various ways other than the approach taken here, but the alternatives explored introduce challenges of their own. Four of these are summarized below.

- If every time t were made a node, this would confront the difficulty that when nodes are dense in time, the sequence of play following a given node may not be well-defined as a function of the strategy profile. See e.g. Stinchcombe (2013), Example 4.2.1.

- If players were permitted to create mid-period nodes arbitrarily many times instead of just once, this would not allow us to solve for within-period behavior (or game-wide behavior, in the finite-horizon cases discussed below) by backward induction.

- The game could be defined so that, at $\xi^i \neq \xi^{-i}, \notin G$, i reveals his own recent spending to $-i$ and lets $-i$ change her plans, without learning $-i$'s recent spending or gaining the ability to change his own plans. This would render $x(h_{|t}, \sigma)$ undefined for some strategy profiles σ and nodes $h_{|t}$ with $t \notin G$. It would thus require a slightly more complex definition of equilibrium. After a

natural re-definition, it would yield the same equilibrium spending behavior as obtains in the game defined above.

- The game could be defined so that players place mid-period nodes without announcing them in advance. However, the natural equilibrium concept is then not SPE but something imposing compatibility between actions and out-of-equilibrium beliefs.

  At $h_{|t}$, even if in equilibrium $-i$ places a node at $\xi^{-i} \in (t, \tau'(t))$, i must plan spending all the way to $\tau'(t)$ in the event that the expected announcement does not arrive. But SPE imposes no restriction on i's spending plans after ξ^{-i}, and unrealistic plans for x^i after ξ^{-i} (such as immediately spending i's entire budget) can in turn deter $-i$ from setting ξ^{-i} later. To rule out these "non-credible threats", we must then ensure that i's spending plans after ξ^{-i} are optimal for i given reasonable beliefs about $-i$'s behavior in the event that $-i$'s expected announcement does not arrive. This too appears to generate similar results but with significantly greater complexity.

Polarization

A polarized [truncated/partial] schedule is a [truncated/partial] schedule x such that, for some time $\tilde{t}$,

i. $x_t^H = 0$ for $t \geq \tilde{t}$ and

ii. $x_t^L = 0$ for $t < \tilde{t}$.

A polarized equilibrium (of game n) is an SPE σ^* such that, for all nodes $h_{|t}$, $x_{[t,\infty)}(h_{|t}, \sigma^*)$ is polarized.

Equilibrium schedules and payoffs

Schedule x is a [polarized] equilibrium schedule if there is a grid sequence and a sequence of [polarized] equilibria $\{\sigma^{(n)}\}$ along the corresponding game sequence such that

i. $x(\sigma^{(n)})$ converges pointwise to x almost everywhere and

ii. $U(x(\sigma^{(n)})) \to U(x)$

as $n \to \infty$.

Note that "polarized equilibrium schedules" are *schedules implemented by polarized equilibria*, rather than merely *equilibrium schedules that are polarized*. These are distinct because an equilibrium may implement a polarized schedule on the equilibrium path (i.e. from the initial node) but not from some off-path nodes.

The game sequences defined above admit many equilibrium schedules. To study the payoff implications of changing the parameters or structure of the games, therefore, it will often be useful to employ a refinement that guarantees uniqueness. As shown in the following subsection, uniqueness is guaranteed if we restrict our attention to equilibrium schedules that resemble the (unique) equilibrium schedules of finite-horizon transformations of the game. This is arguably the behavior it is most natural to take as a baseline, in the absence of coordination on some pattern of promised rewards and punishments.

Define T-horizon game n to be the discrete-time game n defined above, but with the ∞ at the top of the integrals in (117) and (119) replaced with a finite end-time $T > 0$. The definition of T-horizon equilibrium schedule follows immediately.

An equilibrium schedule x is a limit equilibrium schedule if there is a continuum of T-horizon equilibrium schedules $\{x^{[T]}\}$, for $T > 0$, such that $x^{[T]}$ converges pointwise

almost everywhere to x as $T \to \infty$.

A payoff profile U is an equilibrium payoff if there is an equilibrium schedule x with $U(x) = U$.

Existence, uniqueness, and robustness

We impose relatively few restrictions on the grid sequence, and define an equilibrium schedule to be one compatible with any grid sequence, to ensure that results about the uniqueness of a given pattern of equilibrium behavior in continuous time are robust to alternative definitions of equilibrium behavior in continuous time.

The flexible definition of a grid sequence will also let us easily note when an existence result is robust, in that it is compatible with all grid sequences.

3.2 The polarized equilibrium

The open-loop and Stackelberg games have a unique equilibrium, and in each case the equilibrium schedule is polarized. The dynamic game has many equilibrium schedules, as noted in the next subsection. Nevertheless, there is a sense in which polarized behavior is most natural in the dynamic setting as well.

Proposition 5. ***Existence, uniqueness, and Stackelberg-equivalence of limit / polarized equilibrium schedule***

Defining x^* as the Stackelberg schedule of Proposition 4, in the dynamic setting of Section 3.1

a. x^* is the unique limit equilibrium schedule.

b. x^* is the unique polarized equilibrium schedule.

c. x^* is an equilibrium schedule under any grid sequence.

Proof. See Appendix A.5. □

Polarization, and the associated impatient first-mover advantage, are thus not only equilibrium-compatible in an infinite-horizon dynamic setting but uniquely compatible with a natural refinement. No matter how small the difference in discount rates, only polarized behavior is continuous with equilibrium behavior on an ever-lengthening finite horizon.

By contrast, suppose the players have the same discount rate δ, and let $X^{(\delta)}$ denote the δ-optimal collective schedule for the collective budget. Then, trivially, for any horizon (finite or infinite) and any set of grid points, any schedule x with $x^H + x^L = X^{(\delta)}$ is an equilibrium schedule.

3.3 Superior equilibrium payoffs

Proposition 6. ***Superior equilibrium payoffs***

Every feasible payoff profile which is Pareto-superior to $U(x^*)$ is an equilibrium payoff profile.

Proof. See Appendix A.6. □

As noted in Section 1.1, the standard intuition for efficiency results of this kind is that, given perfect monitoring in continuous time, deviating from an efficient equilibrium gives the deviator only an instantaneous benefit (say, by increasing his own consumption while free-riding on other players' public good contributions) but induces a positive-sized punishment (say, by decreasing subsequent contributions). This intuition does not apply straightforwardly to this setting, where there is only a public good, and L's temptation to defect is a temptation to pay an instantaneous cost (by "underspending") for a long-term benefit.

Here, a proof of equilibrium-compatibility for every efficient feasible payoff profile Pareto-superior to $U(x^*)$ can be sketched as follows. A proof for the case of inefficient Pareto-superior payoff profiles is similar.

Given a collective budget B and an efficient collective schedule $X^{(a)}$ that is Pareto-superior to X^*, there is a range of values of b such that both players prefer $X^{(a)}$ to $X^*[b]$, the Stackelberg schedule that would result from b (fixing B). The range may be denoted $[\underline{b}_{X^{(a)}}, \overline{b}_{X^{(a)}}]$. Given the starting budget share for L b_0, and given efficient $X^{(a)}$ with $b_0 \in [\underline{b}_{X^{(a)}}, \overline{b}_{X^{(a)}}]$, can the parties be required to spend so that, for all t, we maintain $b_t \in [\underline{b}_{X^{(a)}_{[t,\infty)}}, \overline{b}_{X^{(a)}_{[t,\infty)}}]$? That is, can we fund $X^{(a)}$ while ensuring that b_t never falls low enough that H would prefer to switch to the Stackelberg schedule implemented given b_t, or rises high enough that L would? If so, there is an SPE in which $X^{(a)}$ is maintained, with deviations punished by reversions to the polarized equilibrium.

L's concern for the shape of the collective schedule in later periods is greater than H's (relative to their concerns for its shape in earlier periods). To be efficient, therefore, X^* must more closely resemble an impatient-optimal schedule early in time and a patient-optimal schedule later in time. So $\underline{b}_{X^{(a)}_{[t,\infty)}}$ and $\overline{b}_{X^{(a)}_{[t,\infty)}}$ increase in t.

An SPE can maintain the value of b around t by requiring the players to contribute to $X^{(a)}_t$ in proportion to their budgets around t. Less obviously, an SPE can induce b_t to rise more quickly than $\overline{b}_{X^{(a)}_{[t,\infty)}}$ around t by requiring H to contribute all of $X^{(a)}_t$. This can be shown by contradiction. Suppose H weakly prefers $X^{(a)}_{[t,\infty)}$ to the Stackelberg schedule that would result from budget-fraction b_t, but, after contributing $x^H_s = X^{(a)}_s$ for $s \in [t, \bar{t})$, disprefers $X^{(a)}_{[\bar{t},\infty)}$ to the Stackelberg schedule with budget-fraction $b_{\bar{t}}$. Then H prefers (i) $x^H_s = X^{(a)}_s, x^L_s = 0$ for $s \in [t, \bar{t})$, followed by a polarized schedule, to (ii) the collective schedule $X^{(a)}_{[t,\infty)}$. But (i) is a polarized

truncated schedule, beginning at t, with the patient budget spent patient-optimally once L begins spending; and H's favorite truncated schedule in this class is the Stackelberg schedule. So H prefers the Stackelberg schedule with budget-fraction b_t to $X^{(a)}_{[t,\infty)}$. This contradicts our assumption.

4 Constraints

As noted in Section 2.2, a static public good game typically exhibits a unique Nash equilibrium, which is "polarized", in the sense that each project is funded entirely by the player(s) who most relatively value it. The prediction of polarization is relatively intuitive in a static context, as there is typically little to prevent polarization in a static public good game. A player is free not to contribute to any given project, and free to contribute as much as she would like. Section 3.2 offers a reason to expect a unique pattern of polarized behavior in a dynamic setting as well, when players have different discount rates and are free to spend as much or as little as they would like at any given time.

But intertemporal polarization may be constrained in two ways. First, a high-discount-rate player may not be able to spend down (the present value of) her budget quickly enough to achieve polarization, e.g. due to a borrowing constraint. Second, a low-discount-rate player may not be able to spend his budget slowly enough due to some sort of saving constraint, such as the legal disbursement minimum to which philanthropic foundations are often subject.

Sections 4.1 and 4.2 introduce these constraints respectively. Unsurprisingly, the constrained game can exhibit what might be called a constrained polarization, with one player bound by the spending minimum or maximum and the other spending more slowly or quickly than they would in isolation to offset the other player's

putative over- or under-spending. In the constrained-polarized setting, however, sufficiently tight constraints of either kind are found to eliminate the first-mover advantage the impatient player enjoys under full polarization in the unconstrained setting.

Section 4.3 then explores the costs that these constraints impose on the players. An asymmetry is found: although spending minima counteract the first-mover advantage, spending minima in this setting can be arbitrarily more costly for the patient, in terms of proportional willingness to pay to remove them, than spending maxima can be for the impatient.

Comparative statics

Given a precisely defined spending constraint, let "constrained game n" be the constrained analog to game n. That is, let it be the game defined in Section 3.1 but with a strategy restriction imposed on the players: a condition that the spending plans assigned to each node must satisfy beyond the budget constraint (118) and right-continuity. The definition of an equilibrium schedule for the constrained game sequence then follows immediately.

To quantify the costs of a given constraint, we will identify a feasible schedule x that is in a natural sense as polarized as is feasible under the constraint, with H spending as quickly and/or L spending as slowly as is feasible (depending on the constraint(s) imposed). We will verify that x is an equilibrium schedule of the constrained game sequence. We will then compare the payoff profiles $U(x)$ and $U(x^*)$. Without picking particular equilibrium schedules in the constrained and unconstrained cases respectively, comparative statics quantifying the impacts of a given constraint are not well defined.

Our motivation for working with an infinite horizon is thus not the standard motivation to explore coordination through repeated interaction. Rather, we will work with an infinite horizon in order to most simply present the results of Section 4.3, which hold in the limit as the time horizon grows large.

Parameter restriction

Throughout this section, we will assume (in addition to parameter restriction (111)) that

$$\gamma \leq 1 \text{ or } \delta^H < \delta^L \frac{\gamma}{\gamma - 1} + r. \tag{120}$$

Note that (120) is (109) with δ^L, δ^H replacing δ, $\tilde{\delta}$ respectively. As in Proposition 2, the motivation for the restriction is to rule out cases in which δ^H-optimal spending yields infinite δ^L-discounted disutility.

4.1 Spending maxima / borrowing constraints

Given i's schedule x^i, let us refer to

$$x_t^i / B_t^i$$

as i's proportional spending rate at t. We will analyze spending maxima that take the form of a constant upper bound $\overline{\alpha}$ on the proportional spending rate that any player can exhibit at any time. Formally, we will restrict the games of Section 3 by requiring that, at each post-announcement node $h_{|t}$, i's spending plan $x^i_{[t,\xi(h_{|t}))}$

satisfy

$$x_s^i \leq B_s^i(x_{|s})\,\overline{\alpha} \tag{121}$$

in addition to the budget constraint (118) and right-continuity.

We will also assume that

$$\overline{\alpha} \in [\alpha^L, \alpha^H]. \tag{122}$$

We will maintain the existence of "announcements" to make the games studied here as comparable as possible to those of Section 3, though they could be eliminated without any impact on the results below.

Motivations

Though spending can of course be constrained in various ways, we will study a proportional spending cap (121) for two reasons.

First, our interest in spending maxima stems in part from the observation that polarization may be limited by the less patient player's borrowing constraints, and a cap on proportional spending is roughly equivalent to a borrowing constraint. Suppose a player has a capital stock at t of $\tilde{B}_t^i$ and a stream of income—e.g. donor contributions—that grows at a constant rate $g < r$, so that inflows at s equal Ce^{gs} for some $C > 0$. The present value (at t) of her budget at t is then

$$B_t^i = \tilde{B}_t^i + \int_t^\infty Ce^{gs}e^{-r(s-t)}ds = \tilde{B}_t^i + \frac{Ce^{gt}}{r-g}.$$

If she cannot borrow against future income, her spending rate at t is unconstrained

as long as $\tilde{B}_t^i > 0$. When $\tilde{B}_t^i = 0$ and her spending is constrained, however, she can spend at an absolute rate of at most Ce^{gt}, and so at a proportional rate of at most $r - g$. And if $\tilde{B}^i > 0$ but i would like to sustain a proportional spending rate greater than $r - g$, then unless she indefinitely refrains from doing so, she eventually exhausts her budget and confronts the proportional spending cap of $r - g$.

Second, upper bounds on proportional spending are precisely analogous to the legal disbursement minima—lower bounds on proportional spending—which are analyzed in Section 4.2 and which this paper is written in part to explore.

Our focus on the $\overline{\alpha} \in [\alpha^L, \alpha^H]$ case (122) also has two motivations.

First, this is the region which most straightforwardly illustrates the implications of constraining the impatient but not the patient from "polarizing". If $\overline{\alpha} < \alpha^L$, so that even L would prefer faster spending than is feasible, then there is little to analyze. The uniquely Pareto-dominant feasible schedule is the one in which $x_t^H/B_t^H = x_t^L/B_t^L = \overline{\alpha}\ \forall t$, as implementable e.g. with the simple SPE in which players spend as quickly as possible at every node. If $\overline{\alpha} > \alpha^H$, on the other hand, then the spending maximum only partially limits polarization by H, and adds kinks to the maximally-polarized equilibrium strategies that complicate the exposition to little benefit. (Recall from Proposition 1 that, in the absence of a more patient funder, H would choose to set $x_t^H/B_t^H = \alpha^H\ \forall t$. In any equilibrium-compatible polarized schedule in which H exhausts her budget at $\overline{t}^*$, $\lim_{t \to \overline{t}^{*-}} x_t^H/B_t^H = \infty$.)

Second, if the rate g at which the players' income streams grow equals the economic growth rate, and if a representative household is not borrowing- or saving-constrained, then the players' proportional spending cap $\overline{\alpha} = r - g$ equals the representative household's optimal proportional spending rate. When the public good under consideration is not a luxury or necessity good, $\overline{\alpha} \in [\alpha^L, \alpha^H]$ then amounts to

the relatively natural assumption that its funders are not all strictly more or strictly less patient than the representative household.

Constrained polarization

Given a spending maximum $\overline{\alpha}$, a constrained-polarized schedule is a schedule x such that $x_t^H = B_t^H \overline{\alpha}$ for all t.

A constrained-polarized equilibrium (of game n) is an SPE σ^* such that, for all nodes $h_{|t}$, $x_{[t,\infty)}(h_{|t}, \sigma^*)$ is constrained-polarized.

A constrained-polarized equilibrium schedule is a schedule x such that there is a grid sequence and a sequence of constrained-polarized equilibria $\{\sigma^{(n)}\}$ across the corresponding game sequence such that $x(\sigma^{(n)})$ converges pointwise to x almost everywhere as $n \to \infty$.

In the absence of a spending maximum, the unique polarized (and unique limit) equilibrium schedule features $x_t^H > B_t^H \alpha^H \geq B_t^H \overline{\alpha}$ whenever $B_t^H > 0$. A constrained-polarized equilibrium schedule is therefore one in which H always chooses a spending rate as close as possible to that she chooses in a polarized equilibrium schedule in the absence of the constraint.

Define the "constrained open-loop game" to be the game in which players simultaneously choose individual schedules x^i, subject to (121), and receive payoffs (117).

Proposition 7. ***Existence and uniqueness of constrained-polarized equilibrium schedule given spending maximum***

Given spending maximum $\overline{\alpha} \in [\alpha^L, \alpha^H]$,

a. There is a unique constrained-polarized equilibrium schedule $\overline{x}$:

$$\overline{x}_t^H = B^H \overline{\alpha} e^{(r-\overline{\alpha})t},$$

$$\overline{x}_t^L = \begin{cases} 0, & t < \overline{t}^*; \\ \left(B^H e^{(r-\overline{\alpha})\overline{t}^*} + B^L e^{r\overline{t}^*}\right)\alpha^L e^{(r-\alpha^L)(t-\overline{t}^*)} - B^H \overline{\alpha} e^{(r-\overline{\alpha})t}, & t \geq \overline{t}^*, \end{cases}$$

where

$$\overline{t}^* \equiv \max\left(0, \ln\left(\frac{B^H}{B^L}\,\frac{\overline{\alpha} - \alpha^L}{\alpha^L}\right)\Big/\overline{\alpha}\right). \tag{123}$$

b. $\overline{x}$ is an equilibrium schedule under any grid sequence.

c. $\overline{x}$ is an equilibrium of the constrained open-loop game.

Proof. See Appendix A.7. □

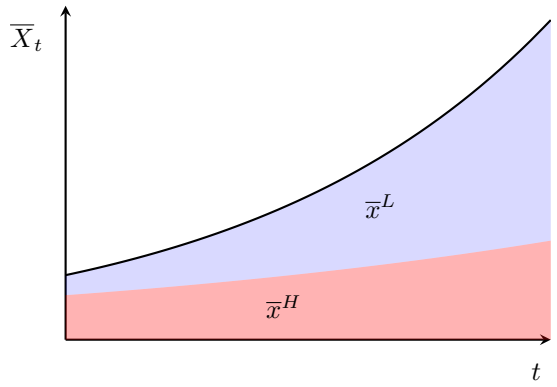


Fig. 3a: Constrained-polarized equilibrium schedule given spending maximum, b large

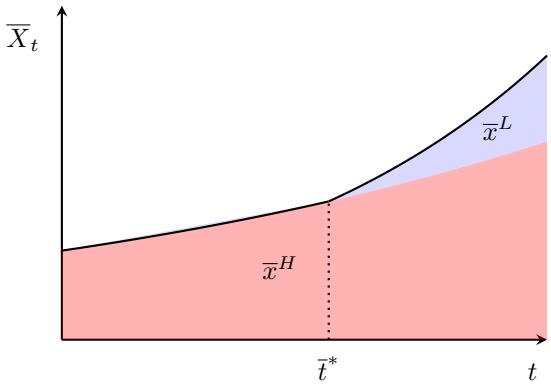


Fig. 3b: Constrained-polarized equilibrium schedule given spending maximum, b small

The spending maximum is binding for H, who spends at proportional rate $\overline{\alpha}$. L then allocates B^L patient-optimally, taking x^H to be independent of his own spending.

If

$$b > 1 - \alpha^L/\overline{\alpha}, \tag{124}$$

then H's fraction of the collective budget is small enough that at time zero her spending rate, $B^H\overline{\alpha}$, is less than the patient-optimal spending rate for the collective budget, $B\alpha^L$. L thus begins spending immediately, and achieves the patient-optimal collective schedule. If the inequality is reversed, H begins by overspending the collective budget, from L's perspective. L compensates by spending nothing until some $\overline{t}^* > 0$, and achieves the patient-optimal collective schedule subsequently.

A behavioral interpretation

Proposition 7 establishes that schedule $\overline{x}$ may obtain given that the players have the preferences from the dynamic setup, given in (117), and strategically interact to maximize them. Only the structure of the games, not the players' preferences, have been modified from Section 3.

Schedule $\overline{x}$ may also be given an alternate interpretation. Suppose that there are no spending constraints, but that H maximizes

$$\int_0^\infty e^{-\delta^H t} u(x_t^H) dt$$

instead of (117). That is, suppose that x^H is a private good for H, rather than X being a public good for both players. Then H chooses $x^H = \overline{x}^H$, in the $\overline{\alpha} = \alpha^H$ case, regardless of σ^L, by Proposition 1. If L still maximizes (117), his best response is $\overline{x}^L$. $\overline{x}$ (with $\overline{\alpha} = \alpha^H$) is thus the schedule that obtains if, in the terminology of Andreoni (1990), H is a warm-glow funder and L is an altruistic funder.

Scenarios in which a less patient player is effectively warm-glow and a more patient player is effectively altruistic may be relevant in practice not only because some funders are in fact warm-glow, but also because a less patient funder may have uncertainty about whether future funding from a patient funder will ever materialize. As noted in the introduction, slow-spending philanthropists sometimes face this very skepticism. The model of Section 3 is then one in which H perfectly anticipates future spending by L, and the model of this section explores the other end of the spectrum, with H expecting no future spending by L at all—at least until he is wealthy enough to implement his favorite allocation of the collective budget.

4.2 Spending minima / saving constraints

We will now explore the implications of a spending minimum $\underline{\alpha}$, in isolation and then in combination with a spending maximum $\overline{\alpha} \geq \underline{\alpha}$. Unlike spending maxima, spending minima are often imposed by law. In the United States, for instance, charitable foundations must disburse at least 5% of their assets each year.

As with the spending maxima of the previous subsection, we will restrict our attention to the $\underline{\alpha} \in [\alpha^L, \alpha^H]$ case. This is again because it is arguably both the most economically interesting case and the most realistic, for roughly the same reasons.

As noted in Section 4.1, borrowing constraints can constrain philanthropic spending on some public good from growing more slowly than the economic growth rate g, which in turn is the optimal spending growth rate from the perspective of a representative household if the public good is not a luxury or a necessity. Disbursement requirements are also sometimes motivated by a desire to prevent foundations from growing quickly enough to accumulate outsize influence.[7] On either grounds, we

[7]In the extreme, see fears that trusts with plans to invest for too long before disbursing would "shatter the nation's financial structure", re-"fashion [the] economy", and ultimately "amount to

might expect to find a legal spending minimum of $r-g$, so that philanthropic spending by all parties grows at a rate no higher than g. In particular, the standard rules of thumb that the real rate of return on capital is 7% per year and the real economic growth rate is 2% per year might be expected to motivate a 5% disbursement minimum, as observed. And again, unless L and H are both unusually patient or impatient, a minimum spending rate of $r-g$ will lie in $[\alpha^L, \alpha^H]$.

Given a spending minimum $\underline{\alpha}$ [and maximum $\overline{\alpha}$], a constrained-polarized schedule is a schedule x such that $x_t^L = B_t^L \underline{\alpha}$ [and $x_t^H = B_t^H \overline{\alpha}$] for all t. Constrained-polarized equilibria and constrained-polarized equilibrium schedules are then defined and motivated as in Section 4.1.

Proposition 8. ***Existence and uniqueness of constrained-polarized equilibrium schedule given spending minimum***

Given spending minimum $\underline{\alpha} \in [\alpha^L, \alpha^H]$,

a. There is a unique constrained-polarized equilibrium schedule $\underline{x}$*:*

$$\underline{x}_t^H = \begin{cases} B^L \alpha^L \left(e^{(\alpha^H - \alpha^L)\underline{t}^* + (r-\alpha^H)t} - e^{(r-\alpha^L)t} \right), & t < \underline{t}^*; \\ 0, & t \geq \underline{t}^*, \end{cases}$$

$$\underline{x}_t^L = B^L \underline{\alpha} e^{(r-\underline{\alpha})t},$$

where $\underline{t}^*$ *uniquely satisfies*

$$\alpha^H / b = \underline{\alpha} e^{(\alpha^H - \underline{\alpha})\underline{t}^*} + (\alpha^H - \underline{\alpha}) e^{-\underline{\alpha}\underline{t}^*}$$

(or equals ∞ *if* $b = 0$*, or if* $\underline{\alpha} = \alpha^H$*).*

all the total value of the world" (O'Kane, 1961).

b. $\underline{x}$ *is an equilibrium schedule under any grid sequence.*

c. $\underline{x}$ *is an equilibrium of the constrained open-loop game.*

Proof. See Appendix A.8. □

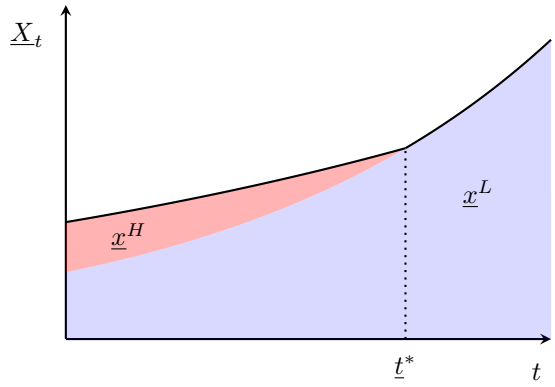


Fig. 4: Constrained-polarized equilibrium schedule given spending minimum

Proposition 9. ***Existence and uniqueness of constrained-polarized limit equilibrium schedule given spending minimum and maximum***

Given spending minimum $\underline{\alpha}$ *and spending maximum* $\overline{\alpha}$ *with* $\alpha^L \leq \underline{\alpha} \leq \overline{\alpha} \leq \alpha^H$,

a. There is a unique limit equilibrium schedule $\overline{\underline{x}}$*:*

$$\overline{\underline{x}}_t^H = B^H \overline{\alpha} e^{(r-\overline{\alpha})t},$$
$$\overline{\underline{x}}_t^L = B^L \underline{\alpha} e^{(r-\underline{\alpha})t}.$$

b. $\overline{\underline{x}}$ *is an equilibrium schedule under any grid sequence.*

c. $\overline{\underline{x}}$ *is an equilibrium of the constrained open-loop game.*

Proof. See Appendix A.9. □

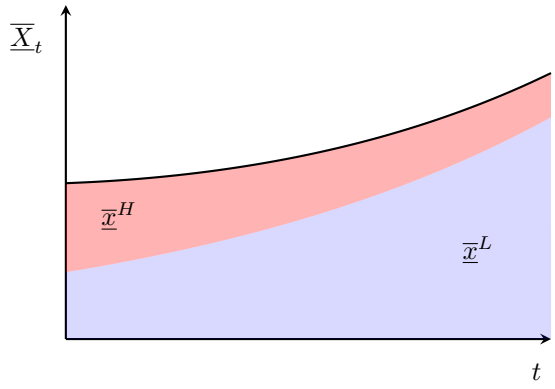


Fig. 5: Constrained-polarized equilibrium schedule given spending minimum and maximum

4.3 The costs of constraints

Despite the symmetry of the spending minima and maxima we have introduced, and the fact that polarized behavior in the unconstrained case gives the impatient player a first-mover advantage, it is the patient player especially who can find a spending constraint costly.

To express this result succinctly, fix r, γ, δ^H, and δ^L satisfying (111) and (120). Then let

$$x[B^H, B^L, \underline{\alpha}, \overline{\alpha}]$$

denote the [constrained-]polarized equilibrium schedule given positive budgets B^L and B^H and a (potentially vacuous) spending minimum $\underline{\alpha}$ and maximum $\overline{\alpha}$ satisfying

$$\underline{\alpha} \in 0 \cup [\alpha^L, \alpha^H], \quad \overline{\alpha} \in [\alpha^L, \alpha^H] \cup \infty, \quad \overline{\alpha} \geq \underline{\alpha}. \tag{125}$$

Finally, given (125), let w^i denote i's proportional willingness to pay to fully relax

the relevant spending constraint:

$$w^H(b, \underline{\alpha}, \overline{\alpha}) \equiv w : U^H(x[1-b, b, \underline{\alpha}, \overline{\alpha}]) = U^H(x[(1-w)(1-b), b, \underline{\alpha}, \infty]),$$
$$w^L(b, \underline{\alpha}, \overline{\alpha}) \equiv w : U^L(x[1-b, b, \underline{\alpha}, \overline{\alpha}]) = U^L(x[1-b, (1-w)b, 0, \overline{\alpha}]).$$

The willingness to pay function $w^i(\cdot)$ is given as a function of L's budget share $b \in (0, 1)$, rather than of B^H and B^L, because holding b fixed, x^*, $\overline{x}$, $\underline{x}$, and $\overline{\underline{x}}$ are linear in B. Also, and relatedly, $U(\cdot)$ is homothetic: given any two schedules to which a player assigns the same payoff, she also assigns the same payoffs to corresponding schedules in which spending at each period is multiplied by a constant $B > 0$. Without loss of generality, therefore, when computing proportional willingness to pay, we can restrict our attention to the $B = 1$ case.

Proposition 10. ***Asymmetric costs of spending constraints***

Given spending constraints $\underline{\alpha}, \overline{\alpha}$ satisfying (125),

a. $\sup_b \ w^H(b, \underline{\alpha}, \overline{\alpha}) < 1$.

b. $\lim_{b \to 0} \ w^L(b, \underline{\alpha}, \overline{\alpha}) = 1$ *if* $\underline{\alpha} \geq \alpha^L$ *and* $\overline{\alpha} > \alpha^L$.

Proof. See Appendix A.10. □

In other words, when the more patient player controls a small proportion of the funding in some domain, a spending minimum is approximately as costly to him as a total expropriation, at least if spending is maximally polarized—even a spending minimum of α^L. This is true whether or not there is also a spending maximum (as long as the spending maximum is not also α^L). A spending maximum is not approximately as costly to the less patient player as a total expropriation, however, and this too is true whether or not there is also a spending minimum.

Recall from Proposition 2 that no result analogous to Proposition 10b obtains in the case of private good provision. This result highlights a welfare-relevant respect in which time preference heterogeneity is especially important in the context of public good provision.

An intuition for the asymmetry in the cost of a spending constraint is clearest when the baseline schedule is constrained in both directions.

Removing a spending maximum does not change x^L in the constrained-polarized equilibrium schedule. When B^H/B is small—i.e. when $b \approx 1$—it simply concentrates H's spending around $t = 0$.

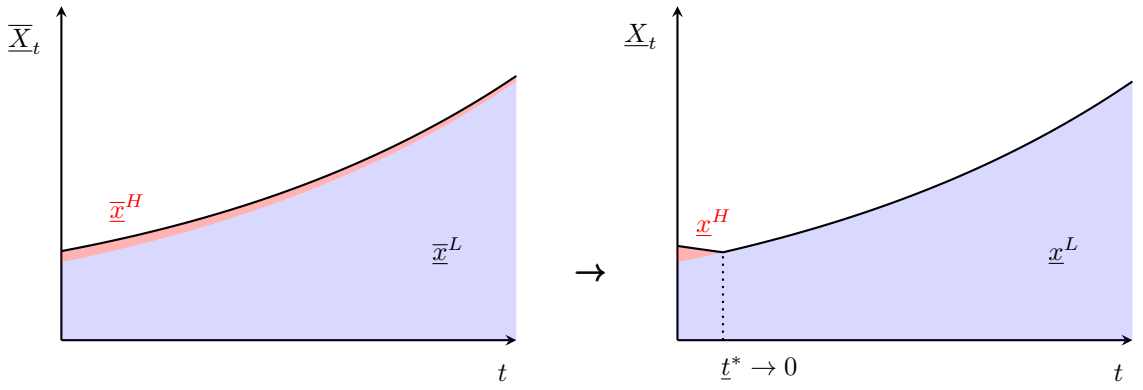


Fig. 6a: Limit equilibrium schedule given binding spending minimum and maximum, $b \to 1$

Fig. 6b: Limit equilibrium schedule given binding spending minimum alone, $b \to 1$

Under both constraints, when $b \approx 1$, the utility that H's budget offers her, above the baseline $U^H(x^L)$ provided by L's spending, is roughly a discounted average of $e^{(r-\delta^H)t}u'(x^L_t)$ across $t \geq 0$ per unit of B^H. When the spending maximum is lifted, the utility that H's budget offers her above the baseline is roughly $u'(x^L_0)$ per unit of B^H. This is a bounded improvement.

With and without the spending minimum, the growth rate of x^H equals $r - \overline{\alpha}$.

Given $\overline{\alpha} > \alpha^L$, is strictly less than the δ^L-optimal spending growth rate of $r - \alpha^L$. So $e^{(r-\delta^L)t}u'(x_t^H)$, the δ^L-discounted marginal utility of an increase in resources allocated to t, grows unboundedly—indeed, exponentially—with t.

Under both constraints, regardless of b, L spreads his spending over time in such a way that most of the present value of his budget is spent before some sufficiently large T. This bounds the additional utility that B^L offers L above the $U^L(x^H)$ baseline.

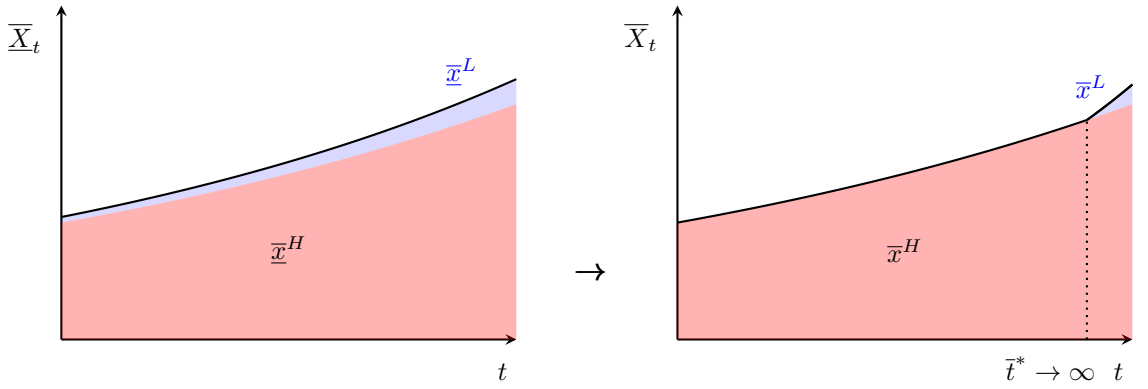


Fig. 7a: Limit equilibrium schedule given binding spending minimum and maximum, $b \to 0$

Fig. 7b: Limit equilibrium schedule given binding spending maximum alone, $b \to 0$

When the spending minimum is lifted, however, the time $\overline{t}^*$ at which L begins spending grows without bound as $b \to 0$. The utility that L's budget offers him above the baseline thus rises to infinity per unit of B^L as $b \to 0$.

Constraints and efficiency

Though efficiency is not the primary focus of this paper, we close this section by noting a way in which constraints on public good spending rates, despite their potentially large costs for one player, can bring about efficiency.

Recall from Proposition 6 that, in the absence of constraints, any efficient payoff Pareto-superior to $U(x^*)$ is an equilibrium payoff. There are also inefficient equilibrium schedules, however, including the polarized equilibrium schedule.

Under a one-sided constraint, the constrained-polarized equilibrium schedule is typically also inefficient. This can be seen from the fact that it typically exhibits a kink in collective spending which both players would prefer to smooth. But if the constraint is tight enough that it implements one player's favorite allocation of the collective budget, the associated schedule is efficient.

Under a two-sided constraint, the constrained-polarized equilibrium schedule features no kink. Instead, (given $\overline{\alpha} > \underline{\alpha}$), the collective proportional spending rate falls smoothly from $b\underline{\alpha} + (1-b)\overline{\alpha}$ to $\underline{\alpha}$ as L gradually comes to possess a larger share of the collective remaining budget. Under certain parameter conditions, this produces an efficient schedule that is not either player's most preferred schedule.

Proposition 11. ***Constraints and efficiency***

Given spending constraints $\underline{\alpha}, \overline{\alpha}$ satisfying (125), $x[B^H, B^L, \underline{\alpha}, \overline{\alpha}]$ is efficient iff

i. $\underline{\alpha} = \alpha^H$*;*

ii. $\overline{\alpha} = \alpha^L$*;*

iii. $\underline{\alpha} = 0$*,* $\overline{\alpha} \geq \alpha^L$*, and* $b > 1 - \alpha^L/\alpha^H$*; or*

iv. $\underline{\alpha} = \alpha^L$*,* $\overline{\alpha} = \alpha^H$*, and* $\gamma = 1$*,*

maximizing U^H given (i), U^L given (ii)-(iii), and $(1-b)U^H + bU^L$ given (iv).

Proof. See Appendix A.11. □

5 Delayed benefits

We have assumed so far that the flow utility the players enjoy at each time depends only on collective spending at that time. This assumption most directly describes the case in which the public good "X" is a perishable consumption good. Often, however, public good expenditures generate benefits after the expenditures occur. Spending on infrastructure, education, research, and environmental preservation is all typically intended to generate a stream of benefits lasting far beyond the day the money is spent. One may therefore wonder whether, in these domains, the basic logic of intertemporal polarization explored in earlier sections is maintained. We will now show that, under broad conditions, it is.

The notion that a more patient player would want to cut such expenditures when a less patient player would want to increase them may be counterintuitive, since calls to increase such expenditures in a political setting are typically made on the basis of their future benefits. But in such a setting, the options at hand are typically (i) to spend on a project with delayed benefits or (ii) to leave the required resources untaxed, in which case at least some of them will be consumed in the short term. By contrast, the setting studied here concerns private funders, such as foundations or charitable trusts, all of whose budgets will be spent on public goods at some time. The options at hand are therefore (i) to spend on a project today or (ii) to invest the entirety of the required resources to a future date and fund more projects then. If an education funder is indifferent between building one school today and two schools in 20 years, it is perhaps not so counterintuitive to suppose that a more patient education funder will prefer the latter.

To formalize this point simply, we will work in discrete time. Let X denote a discrete-

time collective schedule $\{X_t\}_{t\in\mathbb{N}}$. Suppose that each player's utility at t equals $u_t(X)$: an arbitrary time-specific function of collective spending X_s for each $s \in \mathbb{N}$. Player i has a sequence of time preference rates δ_t^i, with $\delta_t^H > \delta_t^L \ \forall t$, so that i's payoff is

$$U^i(X) = \sum_{t=0}^{\infty} e^{-\sum_{q=1}^{t} \delta_q^i} u_t(X). \tag{126}$$

Note that we do not require u_t to be increasing in each X_s: u_t may be constant or decreasing in X_s, locally or globally, for some or all s. In particular, we do not require u_t to be constant in X_s for $s < t$: the plan to invest a unit of resources until t may generate costs or benefits before t, e.g. if the invested resources support environmentally destructive or socially impactful firms.

Because $u_t(\cdot)$ is time-varying, this framework is flexible enough to incorporate the effects of any funders outside the two whose spending constitutes X. For illustration, suppose flow utility at t is the logarithm of spending at t by H, L, and a third party. If the third party exogenously spends one unit at t, then $u_t(X) = \log(X_t + 1)$. If the third party is fully crowded out of spending at t as H and L increase their own spending at t, then $u_t(X) = \log(\max(1, X_t))$.

Like discount rates, we do not require that interest rates be constant or positive. Instead, let $R_0 = 1$, and let $R_t > 0$ denote the cumulative returns to investing from 0 to t.

$\{u_t(\cdot)\}$ satisfies single crossing from X if the function-sequence is differentiable

at X and for each pair of times $\underline{s} < \overline{s}$,

$$\exists \tilde{t} \in \{0, 1, ..., \infty\} \; : \; \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \geq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial u_t(X)}{\partial X_{\overline{s}}} \;\; \forall t < \tilde{t}, \tag{127}$$

$$\frac{\partial u_t(X)}{\partial X_{\underline{s}}} \leq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial u_t(X)}{\partial X_{\overline{s}}} \;\; \forall t > \tilde{t}; \text{ and}$$

$$\exists t \; : \; \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \neq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial u_t(X)}{\partial X_{\overline{s}}}. \tag{128}$$

That is, single crossing obtains from X if a marginal resource allocation to some time $\underline{s}$ produces a stream of undiscounted net benefits that is not equal to the benefit-stream produced by a marginal resource allocation to $\overline{s} > \underline{s}$, but lies weakly above it, lies weakly below it, or crosses it once from above.

The proposition below establishes that single crossing is closely associated with intertemporal polarization. First we will argue for the plausibility of the property across familiar domains in which spending brings delayed benefits.

Suppose $\{u_t(\cdot)\}$ and X are such that marginal education spending at s produces a stream of benefits that begins at C_s per unit of expenditure and then grows exponentially at rate g.[8] Then for any pair $\underline{s} < \overline{s}$, the stream of benefits to allocating marginal resources for education expenditure at $\underline{s}$ lies weakly above that for $\overline{s}$ if $C_{\overline{s}} \leq C_s \, R_s/R_{\overline{s}}$, and the streams cross once (at $\overline{s}$) if $C_{\overline{s}} > C_{\underline{s}} \, R_{\underline{s}}/R_{\overline{s}}$. In either case, single-crossing from X obtains.

Or consider the case of avoiding a catastrophe severe enough to cause human extinction. Because the effects of such a catastrophe would be permanent, the effort to avoid one is sometimes used as the paradigmatic example of a domain in which the patient desire more spending than the impatient. Suppose that, given X, the flow utility associated with the existence of humanity at t is $v_t > 0$ and the probability

[8] This would roughly be implied by the Lucas (1988) model of human capital accumulation, for instance.

of survival to t is p_t, with $p_0 = 1$ and $p_{(\cdot)}$ non-increasing. Then (expected) utility at t is $u_t(X) = p_t v_t$. Suppose also that marginal spending at s lowers the probability of an existential catastrophe at s by C_s per unit spent, given survival to s (resources allocated to s having no effect absent survival to s). Then the stream of expected benefits generated by allocating marginal resources to s is $\{R_s C_s p_t v_t\}_{t \geq s}$. As in the education case, for any pair $\underline{s} < \overline{s}$, the benefit-stream to allocating marginal resources to $\underline{s}$ lies weakly above that for $\overline{s}$ if $C_{\overline{s}} \leq C_s \, R_s / R_{\overline{s}}$, and the streams cross once, at $\overline{s}$, if $C_{\overline{s}} > C_{\underline{s}} \, R_{\underline{s}} / R_{\overline{s}}$.

As in earlier sections, suppose the players begin with budgets $\{B^i\} > 0$, with $B \equiv B^H + B^L$. A collective schedule X is <u>feasible</u> if $\sum_{t=0}^{\infty} X_t / R_t \leq B$, and an individual schedule for i, $x^i \equiv \{x_t^i\}_{t \in \mathbb{N}}$, is feasible if $\sum_{t=0}^{\infty} x_t^i / R_t \leq B^i$.

The <u>generalized open-loop game</u> is the game in which the players simultaneously choose feasible individual schedules and earn payoffs (126).

Proposition 12. ***Polarization and single crossing***

If $\{u_t(\cdot)\}$ *satisfies single crossing*

a. from X*, then given* $\underline{s} < \overline{s}$*,*

$$\frac{\partial U^H(X)}{\partial X_{\overline{s}}} \geq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^H(X)}{\partial X_{\underline{s}}} \implies \frac{\partial U^L(X)}{\partial X_{\overline{s}}} > \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^L(X)}{\partial X_{\overline{s}}},$$

$$\frac{\partial U^L(X)}{\partial X_{\overline{s}}} \leq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^L(X)}{\partial X_{\underline{s}}} \implies \frac{\partial U^H(X)}{\partial X_{\overline{s}}} < \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^H(X)}{\partial X_{\underline{s}}}.$$

b. from all feasible X*, then in any equilibrium* x *of the generalized open-loop game,*

$$\exists \tilde{s} : \; x_s^H = 0 \;\; \forall s > \tilde{s}, \;\; x_s^L = 0 \;\; \forall s < \tilde{s}.$$

Proof. See Appendix A.12. □

It appears that an analysis of dynamic behavior in this generalized setting would be very complex, and it will not be pursued here. Nevertheless, Proposition 12 and the discussion above suggests that, in isolation, the fact that public good spending often produces delayed benefits does not typically alter the basic logic of intertemporal polarization. Patience may motivate postponing one's philanthropy.

6 Conclusion

The economic implications of time preference heterogeneity have been extensively explored in a variety of domains, including social discounting, optimal taxation, and dynamic bargaining. They have to date, however, largely been overlooked in literature on the private provision of public goods. The results presented here begin to fill this gap, and illustrate the importance of building time preference heterogeneity into models of dynamic public good provision going forward.

In dynamic public good provision contexts, time preference heterogeneity appears both even more widespread and even more important than in other contexts.

For the most important dynamic public good games in practice, there is unusually clear empirical and theoretical evidence that their players do not employ common discounting, as noted in Section 1.1.

The results of this paper illustrate that time preference heterogeneity is not just particularly widespread in dynamic public good provision settings but also particularly important. As a comparison between the cases of Section 2.1 and Section 3 illustrates, equilibrium behavior in dynamic public good games with versus without time preference heterogeneity can differ dramatically: a player might spend at a pos-

itive rate at time zero under common discounting but, given a time preference rate only slightly below that of his co-funder, spend nothing for a long while. And even under arbitrarily small time preference differences, spending constraints can have arbitrarily large proportional welfare implications for small patient philanthropists in a large impatient world, in the sense explored in Section 4.3 (though not the reverse). Spending constraints typically have no such extreme implications in the private good context.

The pervasiveness and importance of time preference heterogeneity in public good provision contexts has at least two broad classes of policy implications.

First, it affects the structure of self-enforcing agreements for the provision of public goods among multiple parties who cannot contract, such as national governments. Such agreements can be efficient, as shown in Section 3.3, at least in continuous time and with perfect monitoring. If agreements are designed without accounting for the parties' different discount rates, however, they will generally not be efficient, and may fail to be self-enforcing as intended. The United States's 2017 withdrawal from the Paris Agreement on climate change, for instance, coincided with an explicit impatient shift in Office of Management and Budget policies on social discounting, as requested by the newly elected president.[9] If Americans are simply less patient than their international counterparts, this will in the long run affect democratically-set discounting policy, and punishments (here taking the form of increases in collective emissions) sufficient to deter defections by many countries may be insufficient to deter defections by Americans.

Second, an understanding of the importance of time preference heterogeneity for public good provision might affect the policy conversation around disbursement

[9]See e.g. Li and Pizer (2021).

requirements for philanthropic foundations, trusts, and DAFs. Voters and policy-makers might look more charitably on slow- or non-disbursing charitable investment vehicles once a lack of disbursement is understood not as proof of a tax-avoidance scheme but as, at least possibly, a logical consequence of patient philanthropic planning. Estimates of a patient philanthropist's high ideal willingness to pay to avoid disbursement requirements may also motivate patient philanthropists themselves to fight disbursement requirements more vigorously.

The lessons emphasized here are demonstrated rigorously in the context in which there is a single good and each player's flow utility is an isoelastic function of total flow spending on it. The robustness of the lessons is suggested by the result of Section 5 that intertemporal polarization, at least, is maintained when the framework is extended to allow for much more general preferences.

These results and discussions however only begin to cover the space of possible implications of time preference heterogeneity for public good provision. We have attempted to isolate the implications of time preference heterogeneity by focusing on a model with only two players and a single public good; common knowledge of preferences and of the value of the good; perfect monitoring; and a risk-free AK investment environment. Even in this highly restricted setting, the analysis is incomplete without further work on equilibrium characterization and selection. To quantify the benefits of coordination, for instance, it may be valuable to explore equilibrium selection among efficient SPEs, perhaps using tools from the considerable existing literature on dynamic bargaining under time preference heterogeneity.

More work, both theoretical and empirical, would also be necessary to understand the implications of time preference heterogeneity in real-world public good provision problems which do not conform to the restrictions listed above. For instance, gov-

ernments and philanthropists generally face a variety of public good domains, which produce benefits on different time horizons. Time preference heterogeneity might then be expected to generate polarization both across domains and across time, with patient parties funding an ever larger share of (ever shorter-horizon) domains as they grow to constitute ever larger proportions of the total philanthropic market. Also, given the possibility of risky investments, since the range of projects covered by each player-type will be endogenous to the players' budget shares, a public good funder's own time preferences and those of other philanthropists presumably have implications for the funder's own financial risk-tolerance. Exploring the implications of time preference heterogeneity across the dimensions of both time and domain further would appear to be a valuable topic for future research.

Even the simple model explored here confronts a host of complexities, in part because many useful results from the literature on repeated games (like algorithms to characterize the set of equilibrium payoffs) cannot be used. Extensions along the lines above would doubtless face even more difficulties. Given the pervasive importance of time preference heterogeneity to dynamic public good provision, however, such efforts appear to be worthwhile.

Appendices

A Proofs

A.1 Proof of Proposition 1 and corresponding payoff

Proof of Proposition 1

Given a schedule x, let

$$y_t \equiv e^{-rt}x_t$$

denote the density of resources allocated at time 0 for investment until, followed by spending at, t. The schedule y of present-value spending rates will be called an "allocation", and y_t a "flow allocation". Let

$$v(y_t) \equiv e^{-\delta t}u(e^{rt}y_t)$$

denote the discounted flow utility at t from flow allocation y_t.

Because marginal utility in spending is always positive, y is optimal only if it exhausts the budget.

A budget-exhausting allocation y maximizes utility iff, given y, the marginal flow utility of allocating to each t equals some constant λ almost everywhere:

$$\begin{aligned} v'(y_t) &= \frac{\partial}{\partial y_t}\left[e^{-\delta t}\frac{(e^{rt}y_t)^{1-\gamma}-1}{1-\gamma}\right] = \lambda \ \forall\forall t, \ \ \gamma \neq 1; \\ &= \frac{\partial}{\partial y_t}\left[e^{-\delta t}\ln(e^{rt}y_t)\right] = \lambda \ \forall\forall t, \ \ \gamma = 1 \end{aligned} \tag{129}$$

(where "$\forall\forall$" means "for almost all", i.e. for all but a set of Lebesgue measure zero). If (129) fails, then there are two positive-measure sets of times $\mathcal{T}_1$, $\mathcal{T}_2$ such that $v'(y_t) > v'(y_s)$ for $t \in \mathcal{T}_2$, $s \in \mathcal{T}_1$ (in turn implying $y_s > 0$ for $s \in \mathcal{T}_1$, since $v'(0) = \infty$). y is thus not optimal: a small reallocation from $\mathcal{T}_1$ to $\mathcal{T}_2$ increases U. If y does satisfy (129) almost everywhere, it is optimal, because $U(\cdot)$ is continuous and strictly quasiconcave: $U(a\tilde{x}+(1-a)\tilde{\tilde{x}}) > aU(\tilde{x})+(1-a)U(\tilde{\tilde{x}})$ for any pair of schedules $\tilde{x}$, $\tilde{\tilde{x}}$ that differ on a positive-measure set of times and for which $U(\tilde{x}) = U(\tilde{\tilde{x}})$. If an alternative feasible allocation $\tilde{y}$ would be superior to y, marginal reallocations from y toward $\tilde{y}$ would also be superior, which (129) rules out.

Taking the derivative and rearranging, we have

$$y_t = \lambda^{-\frac{1}{\gamma}} e^{\frac{r-r\gamma-\delta}{\gamma}t} \quad \forall\forall t. \tag{130}$$

Subjecting (130) to the budget constraint (105), we have

$$\int_0^\infty \lambda^{\frac{-1}{\gamma}} e^{\frac{r-r\gamma-\delta}{\gamma}t} dt = B. \tag{131}$$

If $\delta > r(1-\gamma)$, we find

$$\lambda^{-\frac{1}{\gamma}} = B\,\frac{r\gamma - r + \delta}{\gamma}\ \big(= B\alpha\big). \tag{132}$$

Substituting (132) into (130), and recalling that $x_t = e^{rt}y_t$, we have

$$x_t = B\alpha e^{(r-\alpha)t} \quad \forall\forall t. \tag{133}$$

Since the only right-continuous schedule x that satisfies (133) is that which equals the expression above for all t, this is the optimal schedule.

If $\delta \leq r(1-\gamma)$, no λ satisfies (131), so no budget-exhausting y satisfies (129). There is no optimal schedule.

Payoff to following the optimal private schedule

Given $\delta > r(1-\gamma)$, i.e. $\alpha > 0$, the payoff to following schedule (133) equals

$$U = \int_0^\infty e^{-\delta t} u\Big(B\alpha e^{(r-\alpha)t}\Big) dt,$$

which simplifies to

$$U = \begin{cases} \frac{B^{1-\gamma}}{1-\gamma}\alpha^{-\gamma}, & \gamma \neq 1; \\ \\ \frac{\delta \ln(B\delta)+r-\delta}{\delta^2}, & \gamma = 1. \end{cases} \tag{134}$$

A.2 Proof of Proposition 2

The payoff obtained given a loss of fraction w of the budget, without a spending rate requirement, equals (134) with $(1-w)B$ in place of B. Denote the resulting expression $U(B, w)$.

To find the payoff obtained under the spending rate requirement, substitute (107) with $\tilde{\alpha}$ in place of α into (104) and integrate to get

$$U_\delta(B, \tilde{\alpha}) \equiv \begin{cases} \frac{B^{1-\gamma}}{1-\gamma} \frac{\tilde{\alpha}^{1-\gamma}}{\tilde{\alpha}+\delta-\tilde{\delta}}, & \gamma \neq 1; \\ \\ \frac{\delta \ln(B\tilde{\delta})+r-\tilde{\delta}}{\delta^2}, & \gamma = 1, \end{cases} \tag{135}$$

where, given $\tilde{\alpha}$,

$$\tilde{\delta} \equiv \tilde{\alpha}\gamma - r\gamma + r \tag{136}$$

is the time preference rate for which $\tilde{\alpha}$ is the optimal spending rate. Note that the integral is defined iff (109) holds.

Set $U_{\delta}(B, \tilde{\alpha}) = U(B, w)$ and solve for w.

A.3 Proof of Proposition 3

Consider a schedule x such that there are positive-measure sets of times $\mathcal{T}_1$, $\mathcal{T}_2$ with $\sup(\mathcal{T}_1) < \inf(\mathcal{T}_2)$, $x_t^L > 0 \; \forall t \in \mathcal{T}_1$, and $x_t^H > 0 \; \forall t \in \mathcal{T}_2$. We begin by showing that x is not an equilibrium.

If $e^{(r-\delta^L)t}(x_t^H + x_t^L)^{-\gamma}$ is not constant almost everywhere in $\mathcal{T}_1$, L prefers to reallocate his spending within $\mathcal{T}_1$. Likewise, if $e^{(r-\delta^H)t}(x_t^H + x_t^L)^{-\gamma}$ is not constant almost everywhere in $\mathcal{T}_2$, H prefers to reallocate her spending within $\mathcal{T}_2$. In either case, x is not an equilibrium.

Suppose therefore that

$$e^{(r-\delta^L)t}(x_t^H + x_t^L)^{-\gamma} = \lambda^L \;\; \forall \forall t \in \mathcal{T}_1, \tag{137}$$

$$e^{(r-\delta^H)t}(x_t^H + x_t^L)^{-\gamma} = \lambda^H \;\; \forall \forall t \in \mathcal{T}_2. \tag{138}$$

If x is an equilibrium, we must have

$$e^{(r-\delta^L)t}(x_t^H + x_t^L)^{-\gamma} \leq \lambda^L \;\; \forall \forall t \in \mathcal{T}_2. \tag{139}$$

Otherwise L prefers to marginally reallocate his spending from $\mathcal{T}_1$ to $\mathcal{T}_2$. Likewise,

we must have

$$e^{(r-\delta^H)t}(x_t^H + x_t^L)^{-\gamma} \leq \lambda^H \quad \forall\forall t \in \mathcal{T}_1, \tag{140}$$

or else H prefers to marginally reallocate her spending from $\mathcal{T}_2$ to $\mathcal{T}_1$. Multiplying both sides of (139) by $e^{(\delta^L-\delta^H)t}$ and both sides of (140) by $e^{(\delta^H-\delta^L)t}$, substituting for the left-hand sides by (138) and (137), and rearranging, we have

$$\begin{aligned} e^{(\delta^H-\delta^L)t} &\leq \lambda^L/\lambda^H \quad \forall\forall t \in \mathcal{T}_2, \\ e^{(\delta^H-\delta^L)t} &\geq \lambda^L/\lambda^H \quad \forall\forall t \in \mathcal{T}_1. \end{aligned}$$

Since $\delta_H > \delta_L$ and $\mathcal{T}_2$ follows $\mathcal{T}_1$, these inequalities cannot simultaneously obtain. This completes the proof that x is not an equilibrium.

It follows that, in any equilibrium, there is a time t^o such that H is the sole funder $\forall\forall t < t^o$ and L is the sole funder $\forall\forall t > t^o$.

As in Proposition 1, each player i does best to spend such that the value of marginal allocations is equal across almost all times at which she spends. This implies that if x is an equilibrium,

$$\begin{aligned} x_t^H &= \Lambda^H e^{(r-\alpha^H)t} \quad \forall\forall t < t^o, \qquad (141) \\ &= 0 \qquad\qquad\quad \forall\forall t > t^o; \\ x_t^L &= 0 \qquad\qquad\quad \forall\forall t < t^o, \qquad (142) \\ &= \Lambda^L e^{(r-\alpha^L)t}, \quad \forall\forall t > t^o, \end{aligned}$$

for some constants Λ^H, Λ^L.

From the budget constraints

$$\int_0^{t^o} \Lambda^H e^{-\alpha^H t} dt = B^H;$$
$$\int_{t^o}^{\infty} \Lambda^L e^{-\alpha^L t} dt = B^L,$$

we then have

$$\Lambda^H = B^H \alpha^H \Big(1 - e^{-\alpha^H t^o}\Big)^{-1}; \tag{143}$$
$$\Lambda^L = B^L \alpha^L e^{-\alpha^L t^o}. \tag{144}$$

An equilibrium collective schedule cannot exhibit a jump discontinuity at its t^o. If collective spending rose discontinuously at t^o, L would prefer a reallocation from just after t^o to just before. Likewise, if it fell discontinuously, H would prefer a reallocation from just before t^o to just after. We therefore have

$$\Lambda^H e^{(r-\alpha^H)t^o} = \Lambda^L e^{(r-\alpha^L)t^o}$$
$$\implies t^o = \frac{1}{\alpha^H - \alpha^L} \ln\Big(\frac{\Lambda^H}{\Lambda^L}\Big). \tag{145}$$

Substituting (145) into (143)–(144) and simplifying, we get

$$\Lambda^H = B^H \alpha^H + B^L \alpha^L; \tag{146}$$
$$\Lambda^L = B^L \alpha^L \Big(1 + \frac{B^H \alpha^H}{B^L \alpha^L}\Big)^{\frac{\alpha^L}{\alpha^H}}. \tag{147}$$

Substituting (146)–(147) into (141)–(142) and (145) produces final expressions for x^H, x^L, and t^o.

We have shown that, if an equilibrium exists, it must take the form above. To

verify that the profile (x^H, x^L) above is in fact an equilibrium, observe that, by construction, from (x^H, x^L), neither player prefers marginal reallocations across times during which he spends, or from times during which he spends to times during which he does not. Because $U^i(\cdot)$ is strictly quasiconcave for both i, neither player prefers a non-marginal alternative feasible $\tilde{x}^i$, fixing $\tilde{x}^{-i}$, either. x^L and x^H are mutual best responses.

A.4 Proof of Proposition 4

Preliminaries

Let $x^L(x^H)$ and $y^L(x^H)$ denote L's best-response individual schedule and allocation, respectively, to H's individual schedule x^H. The strict quasiconcavity of $U(\cdot)$ guarantees that if a best response exists, it is unique (up to measure-zero deviations). We will construct the unique form that an equilibrium $(x^H, x^L(x^H))$ must take if one exists and then show that the strategies are indeed mutual best responses.

Given x^H, let

$$\mathcal{T}^L(x^H) \equiv \{t : x_t^L(x^H) > 0\}, \quad \mathcal{T}^H(x^H) \equiv \{t : x_t^H > 0\}.$$

Note that the budget constraints

$$\int_{\mathcal{T}^L(x^H)} y_t^L(x^H)dt \leq B^L, \quad \int_{\mathcal{T}^H(x^H)} y_t^H dt \leq B^H$$

imply that $\mathcal{T}^L(x^H)$ and $\mathcal{T}^H(x^H)$, and thus their intersection, are measurable. Let

$$\mathcal{B}(x^H) \equiv \int_{\mathcal{T}^L(x^H)} y_t^L(x^H) \ \ (\leq B^H)$$

denote the budget that H allocates to $\mathcal{T}^L(x^H)$.

$y^L(x^H)$ must set the δ^L-discounted marginal utility of flow allocations equal $\forall\forall t \in \mathcal{T}^L(x^H)$. As in Proposition 1, this holds iff

$$x_t^H + x_t^L(x^H) = \Lambda^L(x^H)e^{(r-\alpha^L)t} \ \ \forall\forall t \in \mathcal{T}^L(x^H), \tag{148}$$

for the constant $\Lambda^L(x^H)$ satisfies the budget constraint

$$\int_{\mathcal{T}^L(x^H)} e^{-rt}(x_t^H + x_t^L(x^H))dt = \mathcal{B}(x^H) + B^L. \tag{149}$$

The resulting $\{x_t^H + x_t^L(x^H)\}_{t\in\mathcal{T}^L(x^H)}$ will implement the unique δ^L-optimal allocation of $\mathcal{B}(x^H) + B^L$ across $\mathcal{T}^L(x^H)$.

Let $\mathcal{T} \subset \mathcal{T}^L(x^H)$ denote a set of times such that

$$\int_{\mathcal{T}} e^{-rt}(x_t^H + X_t^L(x^H))dt = \mathcal{B}(x^H).$$

Such $\mathcal{T}$ must exist, by the continuity of total resource allocation with respect to time. If $\mathcal{B}(x^H) > 0$ any such $\mathcal{T}$ must have positive measure.

Suppose $\mathcal{B}(x^H) > 0$, and consider an $\tilde{x}^H$ such that

$$\tilde{x}_t^H = \begin{cases} x_t^H, & t \notin \mathcal{T}^L(x^H); \\ x_t^H + x_t^L(x^H), & t \in \mathcal{T}; \\ 0, & t \in \mathcal{T}^L(x^H)\backslash\mathcal{T}. \end{cases}$$

L will then still be able to achieve a collective schedule of $x_t^H + x_t^L(x^H)$, by adopting

individual schedule $\tilde{x}^L$ with

$$\tilde{x}_t^L = \begin{cases} 0, & t \notin \mathcal{T}^L(x^H) \backslash \mathcal{T}; \\ x_t^H + x_t^L(x^H), & t \in \mathcal{T}^L(x^H) \backslash \mathcal{T}. \end{cases}$$

$(\tilde{x}^H, \tilde{x}^L)$ still implements the unique δ^L-optimal allocation of resources $\mathcal{B}(x^H) + B^L$ across $\mathcal{T}^L(x^H)$. Furthermore, because $\tilde{x}_t^H = x_t^H \ \forall t \notin \mathcal{T}^L(x^H)$, L still weakly prefers marginal spending within $\mathcal{T}^L(x^H)$ to marginal spending outside $\mathcal{T}^L(x^H)$. Given $\tilde{x}^H$, therefore, L does indeed best respond with $\tilde{x}^L$; $\tilde{x}^L = x^L(\tilde{x}^H)$.

$\tilde{x}^H$ thus induces the same collective schedule as x^H. However, $\mathcal{B}(\tilde{x}^H) = 0$. We have found that, for any feasible x^H, there is a feasible $\tilde{x}^H$ such that

$$\mathcal{B}(\tilde{x}^H) = 0 \tag{150}$$

and H is indifferent between x^H and $\tilde{x}^H$ given L's best response.

Given a feasible x^H satisfying (150), let

$$\hat{\mathcal{T}}(x^H) \equiv \{t : x_t^H > \Lambda^L(x^H) e^{(r-\alpha^H)t}\},$$
$$\hat{\mathcal{B}}(x^H) \equiv \int_{\hat{\mathcal{T}}(x^H)} y_t^H dt.$$

Observe by the reasoning of Proposition 1 that there is a (near-)unique and δ^H-optimal feasible $\tilde{x}^H$, under the condition $\mathcal{T}^H(\tilde{x}^H) = \mathcal{T}^H(x^H)$, such that

$$\tilde{x}_t^H = \begin{cases} 0, & t \notin \mathcal{T}^H(x^H); \\ \max(\Lambda^H(\tilde{x}^H) e^{(r-\alpha^H)t}, \Lambda^L(\tilde{x}^H) e^{(r-\alpha^L)t}), & t \in \mathcal{T}^H(x^H), \end{cases} \tag{151}$$

where $\Lambda^L(\tilde{x}^H) = \Lambda^L(x^H)$ is given by (148)–(149) and

$$\Lambda^H(\tilde{x}^H) \equiv \begin{cases} \Lambda^H : \int_{\hat{\mathcal{T}}(x^H)} \Lambda^H e^{-\alpha^H t} dt = \hat{\mathcal{B}}(x^H) & \\ \qquad \Longrightarrow \ \hat{\mathcal{B}}(x^H) \big/ \int_{\hat{\mathcal{T}}(x^H)} e^{-\alpha^H t} dt, & \hat{\mathcal{B}}(x^H) > 0; \\ \Lambda^L(\tilde{x}^H), & \hat{\mathcal{B}}(x^H) = 0. \end{cases}$$

That is, H always does best to allocate δ^H-optimally across the periods during which she spends, at least subject to the constraint that such a shift does not render her spending at a time low enough that L responds by altering his own schedule.

Consider a feasible allocation y^H (and corresponding x^H) satisfying (150) and (151). We must have

$$y_t^H = \begin{cases} \Lambda^H(x^H) e^{-\alpha^H t}, & t \in \mathcal{T}^H(x^H) \cap [0, \overline{q}]; \\ \Lambda^L(x^H) e^{-\alpha^L t}, & t \in \mathcal{T}^H(x^H) \cap [\overline{q}, \infty); \end{cases} \tag{152}$$

$$y_t^L(x^H) = \quad \Lambda^L(x^H) e^{-\alpha^L t}, \quad t \in \mathcal{T}^L(x^H),$$

where

$$\overline{q}(x^H) \equiv \ln\Big(\frac{\Lambda^L(x^H)}{\Lambda^H(x^H)}\Big) \big/ (\alpha^L - \alpha^H)$$

denotes the time q at which $\Lambda^H(x^H) e^{-\alpha^H q} = \Lambda^L(x^H) e^{-\alpha^L q}$.

Define

$$\underline{Q}(q;x^H) \equiv \int_{\mathcal{T}^L(x^H)\cap[0,q)} \Lambda^L(x^H)e^{-\alpha^L t}, \tag{153}$$
$$\overline{Q}(q;x^H) \equiv \int_{\mathcal{T}^H(x^H)\cap[\max(\overline{q},q),\infty)} \Lambda^L(x^H)e^{-\alpha^L t}.$$

Since $\underline{Q}(q;x^H)$ weakly and continuously increases in q from zero to B^L, and $\overline{Q}(q;x^H)$ weakly and continuously decreases in q from a nonnegative value to zero, there exists a (not necessarily unique) q^* such that

$$\underline{Q}(q^*;x^H) = \overline{Q}(q^*;x^H).$$

In equilibrium, q^* is a time such that the amount (in present-value terms) that L allocates δ^L-optimally before q^* equals the amount that H allocates δ^L-optimally after q^*.

Choosing some such q^*, now consider the allocation

$$\tilde{y}_t^H = \begin{cases} \Lambda^L(x^H)e^{-\alpha^L t}, & t \in \mathcal{T}^L(x^H)\cap[0,q^*); \\ 0, & t \in \mathcal{T}^H(x^H)\cap[\max(\overline{q}(x^H),q^*),\infty); \\ y_t^H, & \text{elsewhere.} \end{cases}$$

Observe that $\Lambda^L(\tilde{x}^H) = \Lambda^L(x^H)$ and $\Lambda^H(\tilde{x}^H) = \Lambda^H(x^H)$, and thus that $\overline{q}(\tilde{x}^H) = \overline{q}(x^H)$.

It follows from (153) that L's (near-)unique best response to a shift from y^H to $\tilde{y}^H$ is to shift his spending from $\mathcal{T}^L(x^H)\cap[0,q^*)$ to $\mathcal{T}^H(x^H)\cap[\max(\overline{q}(x^H),q^*),\infty)$, leaving his spending elsewhere unchanged; this alone maintains (148). It likewise follows from (153) that $\tilde{y}^H$ is affordable for H, and that it induces the same collective

allocation as y^H. Finally, $\sup(\mathcal{T}^H(\tilde{x}^H)) \leq \max(\overline{q}(\tilde{x}^H), q^*)$, so

$$\begin{aligned} T(\tilde{x}^H) &\equiv [\inf(\mathcal{T}^L(\tilde{x}^H)), \sup(\mathcal{T}^H(\tilde{x}^H))] \text{ is bounded and} \\ \tilde{x}^H_t &= \Lambda^H(\tilde{x}^H) e^{(r-\alpha^H)t} \ \ \forall t \in T(\tilde{x}^H), \end{aligned} \tag{154}$$

with the second observation following from (151)–(152) and the fact that $\overline{q}(\tilde{x}^H) = \overline{q}(x^H)$. Note that $\inf(\mathcal{T}^L(x^L(\tilde{x}^H))) \geq q^*$, so $T(\tilde{x}^H)$ is of positive measure only if $\max(T(\tilde{x}^H)) = \overline{q}(\tilde{x}^H)$.

We have now shown that, for any feasible x^H satisfying (150) and (151), there is a feasible $\tilde{x}^H$ satisfying (150) and (154) such that H is indifferent between x^H and $\tilde{x}^H$ given L's best response.

Without loss of generality, let us define $x^L(\cdot)$ such that

$$\int_{T(x^H)} e^{-rt} x^L_t(x^H) dt > 0 \ \ \forall x^H : \inf(\mathcal{T}^L(x^H)) < \sup(\mathcal{T}^H(x^H)). \tag{155}$$

That is, let us not say that L ever responds to x^H by spending at a measure-zero set of times before he begins to allocate positive resources.

If

$$\int_{T(x^H)} e^{-rt} x^H_t dt = 0, \tag{156}$$

x^H is a measure-zero deviation from an alternative strategy for H, $\tilde{x}^H$ with

$$\tilde{x}^H_t = \begin{cases} x^H_t, & t < \inf(\mathcal{T}^L(x^H)); \\ 0, & t \geq \inf(\mathcal{T}^L(x^H)), \end{cases} \tag{157}$$

such that the schedule $(\tilde{x}^H, x^L(\tilde{x}^H))$ is polarized in the sense that

$$\exists \tilde{t}(\tilde{x}^H) \ : \ x_t^L(\tilde{x}^H) = 0 \ \ \forall t < \tilde{t}(\tilde{x}^H), \quad \tilde{x}_t^H = 0 \ \ \forall t \geq \tilde{t}(\tilde{x}^H). \tag{158}$$

Of course, $x^L(\tilde{x}^H) = x^L(x^H)$ up to measure-zero deviations and, given these best responses, H is indifferent between x^H and $\tilde{x}^H$.

The next step of the proof is to show that for any feasible x^H satisfying (150) and (154) but not (156), there is a feasible $\tilde{x}^H$, which H in equilibrium strictly prefers to x^H, satisfying all three conditions. It will then follow that, in equilibrium, x^H must be at most a zero-measure deviation from an $\tilde{x}^H$ satisfying (158). We will then be able to determine the (near-)uniquely optimal x^H in this class, and thus the equilibrium as a whole.

From here we will proceed differently depending on the value of γ.

$\gamma > 1$ case

Consider a feasible allocation y^H (and corresponding x^H) satisfying (150) and (154) but not (156). Recall that this implies that $\max(T(x^H)) \leq \overline{q}(x^H)$.

Define

$$\underline{Z}(z; x^H) \equiv \int_{\mathcal{T}^L(x^H) \cap [0,z)} e^{-\alpha^L t} dt, \tag{159}$$

$$\overline{Z}(z; x^H) \equiv \int_{\mathcal{T}^H(x^H) \cap [z, \overline{q}(x^H))} e^{-\alpha^L t} dt. \tag{160}$$

Since $\overline{Z}(z; x^H) - \underline{Z}(z; x^H)$ strictly and continuously decreases in z from a positive value at $z = 0$ to a negative (by (155)) value at $z = \overline{q}(x^H)$, there is a unique

$z^* \in (0, \overline{q}(x^H))$ such that $\underline{Z}(z^*; x^H) = \overline{Z}(z^*; x^H)$. Let

$$\begin{aligned} \underline{T}(x^H) &\equiv \mathcal{T}^L(x^H) \cap [0, z^*), \\ \overline{T}(x^H) &\equiv \mathcal{T}^H(x^H) \cap [z^*, \overline{q}(x^H)). \end{aligned} \tag{161}$$

It follows from the fact that $\sup(\mathcal{T}^H(x^H)) \leq \overline{q}(x^H)$ that $y_t^H \geq \Lambda^L(x^H)e^{-\alpha^L t}$ $\forall t \in \mathcal{T}^H(x^H)$.

Choose $\epsilon > 0$, and partition $\overline{T}(x^H)$ into (not necessarily nonempty) elements

$$\overline{T}_{j,\epsilon}(x^H) \equiv \overline{T}(x^H) \cap [z^* + j - \epsilon, z^* + j) \tag{162}$$

for all $j \in \epsilon\mathbb{N}$, where $\epsilon\mathbb{N}$ denotes $\{\epsilon, 2\epsilon, ...\}$. Also, define

$$\underline{t}_j(x^H) \equiv \min\Big\{q : \int_{\underline{T}(x^H)\cap[0,q)} e^{-\alpha^L t}dt = \int_{\overline{T}(x^H)\cap[z^*,z^*+j)} e^{-\alpha^L t}dt\Big\}, \tag{163}$$

$$\underline{T}_{j,\epsilon}(x^H) \equiv \underline{T}(x^H) \cap [\underline{t}_{j-\epsilon}(x^H), \underline{t}_j(x^H)), \tag{164}$$

$$j(t, \epsilon; x^H) \equiv j \ : \ t \in \underline{T}_{j,\epsilon}(x^H) \cup \overline{T}_{j,\epsilon}(x^H). \tag{165}$$

That is, $j(t, \epsilon; x^H)$ is the element of $\epsilon\mathbb{N}$ such that t lies in the element of the partitioned $\overline{T}(x^H)$ whose supremum is j, or the corresponding element of the partitioned $\underline{T}(x^H)$.

Let

$$S(x^H) \equiv \Big\{s \in \overline{T}(x^H) \ : \ \lim_{\epsilon\to 0}\big(\underline{t}_{j(s,\epsilon;x^H)}(x^H) - \underline{t}_{j(s,\epsilon;x^H)-\epsilon}(x^H)\big) > 0\Big\}. \tag{166}$$

It follows from (163) that, for all $s \in S(x^H)$,

$$\exists \phi > 0 : \int_{\underline{T}(x^H) \cap [0, \underline{t}_s(x^H))} e^{-\alpha^L t} dt = \int_{\underline{T}(x^H) \cap [0, \underline{t}_s(x^H) + \phi)} e^{-\alpha^L t} dt = \int_{\overline{T}(x^H) \cap [z^*, s)} e^{-\alpha^L t} dt. \tag{167}$$

Given $s \in S(x^H)$, let ϕ_s denote the supremum ϕ satisfying (167). Thus, for each $s \in S(x^H)$, there is a maximal subinterval $\Phi(s) \equiv [\underline{t}_s(x^H), \phi_s)$ of $[0, z^*)$ such that $\mu(\underline{T}(x^H) \cap \Phi(s)) = 0$. Because any interval can be partitioned into at most countably many subintervals, there are at most countably many such Φ. Furthermore, for each Φ, we must have

$$\mu\big(\{s : \Phi(s) = \Phi\}\big) = 0;$$

otherwise the integral on the right-hand side of (167) could not be equal for all such s. Therefore $\mu(S(x^H)) = 0$.

It follows that, given any feasible x^H satisfying (150) and (154), there is a corresponding feasible $\tilde{x}^H$ with

$$\tilde{x}^H_t = \begin{cases} x^H_t, & t \notin S(x^H); \\ 0, & t \in S(x^H), \end{cases}$$

also satisfying (150) and (154), but for which we also have

$$S(\tilde{x}^H) = \emptyset. \tag{168}$$

Of course $x^L(\tilde{x}^H) = x^L(x^H)$ almost everywhere, and H is indifferent between x^H and $\tilde{x}^H$ given L's best responses.

Consider an allocation y^H (and corresponding x^H) satisfying (150), (154), and (168) but not (156). Then, given a choice of interval-length ϵ, consider the allocation $\tilde{y}^H(\epsilon)$ (and corresponding $\tilde{x}^H(\epsilon)$) with

$$\tilde{y}_t^H(\epsilon) = \begin{cases} y_t^H, & t \notin \overline{T}(x^H) \cup \underline{T}(x^H); \\ \Lambda^H(x^H) e^{(\alpha^L - \alpha^H)(z^* + j(t,\epsilon;x^H)) - \alpha^L(\underline{t}_{j(t,\epsilon;x^H)}(x^H) + \epsilon)}, & t \in \underline{T}(x^H); \\ 0, & t \in \overline{T}(x^H). \end{cases}$$

Note that $\tilde{x}^H(\epsilon)$ satisfies (156) for all ϵ.

To demonstrate that $\tilde{y}^H(\epsilon)$ is feasible, let us show that its allocation to each $\underline{T}_{j,\epsilon}(x^H)$ is weakly (and indeed strictly) less than y^H's allocation to the corresponding $\overline{T}_{j,\epsilon}(x^H)$. From (163) and (164), we have

$$\int_{\underline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L t} dt = \int_{\overline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L t} dt. \tag{169}$$

Observe that $t < \underline{t}_j(x^H)\ \forall t \in \underline{T}_{j,\epsilon}(x^H)$ and $t \geq z^* + j - \epsilon\ \forall t \in \overline{T}_{j,\epsilon}(x^H)$. Also,

$$y_t^H \geq \Lambda^H(x^H) e^{-\alpha^H(z^*+j)} \ \ \forall t \in \overline{T}_{j,\epsilon}(x^H).$$

Thus (169) gives

$$\int_{\underline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L \underline{t}_j(x^H)} dt \leq \int_{\overline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L(z^*+j-\epsilon)} dt \tag{170}$$

$$\implies \int_{\underline{T}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{(\alpha^L - \alpha^H)(z^*+j) - \alpha^L(\underline{t}_j(x^H)+\epsilon)} dt \leq \int_{\overline{T}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{-\alpha^H(z^*+j)} dt$$

$$< \int_{\overline{T}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{-\alpha^H t} dt. \tag{171}$$

Summing across $j \in \epsilon\mathbb{N}$, it follows that, since y^H is feasible, $\tilde{y}^H(\epsilon)$ is also feasible.

Now let us show that, for sufficiently small ϵ, H strictly prefers $\tilde{y}^H(\epsilon)$ to y^H given L's best responses.

First, for any ϵ, we can decompose the move from y^H to $\tilde{y}^H(\epsilon)$ into a sequence of shifts from $\{y_t^H\}$ to $\{\tilde{y}_t^H(\epsilon)\}$ for $t \in \underline{T}_{j,\epsilon}(x^H) \cup \overline{T}_{j,\epsilon}(x^H)$ for each j, with the original allocation maintained elsewhere. That is, H can shift spending back from $\overline{T}(x^H)$ to $\underline{T}(x^H)$ by shifting spending back from $\overline{T}_{j,\epsilon}(x^H)$ to $\underline{T}_{j,\epsilon}(x^H)$ for each j. Each shift will be affordable, as shown by (171). Furthermore, from (163) and (164) we see that, in equilibrium, L will (up to measure-zero deviations) respond to each shift by shifting his own allocated funds from $\underline{T}_{j,\epsilon}(x^H)$ to $\overline{T}_{j,\epsilon}(x^H)$, leaving his spending elsewhere unchanged; this alone maintains condition (148) for some Λ^L. Finally, observe that the shift must increase flow spending and thus utility throughout $\underline{T}_{j,\epsilon}(x^H)$ and decrease it throughout $\overline{T}_{j,\epsilon}(x^H)$.

H's net utility gain from shift j is thus bounded below by

$$\frac{1}{1-\gamma}\Bigg[\int_{\underline{T}_{j,\epsilon}(x^H)} e^{-\delta^H \underline{t}_j(x^H)}\Big(\big(\Lambda^H(x^H)e^{(\alpha^L-\alpha^H)(z^*+j)-\alpha^L(\underline{t}_j(x^H)+\epsilon)+r\underline{t}_j(x^H)}\big)^{1-\gamma} - \big(\Lambda^L(x^H)e^{(r-\alpha^L)\underline{t}_{j-\epsilon}(x^H)}\big)^{1-\gamma}\Big)dt \tag{172}$$

$$+\int_{\overline{T}_{j,\epsilon}(x^H)} e^{-\delta^H(z^*+j-\epsilon)}\Big(\big(\Lambda^L(x^H)e^{(r-\alpha^L)(z^*+j)}\big)^{1-\gamma} - \big(\Lambda^H(x^H)e^{(r-\alpha^H)(z^*+j-\epsilon)}\big)^{1-\gamma}\Big)dt\Bigg].$$

From (169), the fact that $t \geq \underline{t}_{j-\epsilon}(x^H)\ \forall t \in \underline{T}_{j,\epsilon}(x^H)$, and the fact that $t < z^*+j\ \forall t \in \overline{T}_{j,\epsilon}(x^H)$, we have

$$\int_{\overline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L(z^*+j)}dt \leq \int_{\underline{T}_{j,\epsilon}(x^H)} e^{-\alpha^L\underline{t}_{j-\epsilon}(x^H)}dt. \tag{173}$$

After rearranging (172) and making the substitution from (173), we see that H's net

utility gain from shift j is further bounded below by

$$\frac{1}{\gamma-1}\big(\Lambda^L(x^H)e^{(r-\alpha^L)\underline{t}_{j-\epsilon}(x^H)}\big)^{1-\gamma}\int_{\overline{T}_{j,\epsilon}(x^H)}e^{-\alpha^L(z^*+j)}dt \tag{174}$$
$$\Big[\Big(1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j)+(r-\alpha^L)(\underline{t}_j(x^H)-\underline{t}_{j-\epsilon}(x^H))-\alpha^L\epsilon}\Big)^{1-\gamma}\Big)e^{\alpha^L\underline{t}_{j-\epsilon}(x^H)-\delta^H\underline{t}_j(x^H)}$$
$$-\Big(1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j)-(r-\alpha^H)\epsilon}\Big)^{1-\gamma}\Big)$$
$$e^{(1-\gamma)(r-\alpha^L)(z^*+j-\underline{t}_{j-\epsilon}(x^H))-\delta^H(z^*+j-\epsilon)+\alpha^L(z^*+j)}\Big].$$

By (166) and (168),

$$\lim_{\epsilon\to0}\underline{t}_{j(t,\epsilon;x^H)}(x^H)=\lim_{\epsilon\to0}\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H)=\underline{t}_{t-z^*}(x^H)\ \ \forall t\in\overline{T}(x^H). \tag{175}$$

Also, by (162) and (165),

$$\lim_{\epsilon\to0}j(t,\epsilon;x^H)=t-z^*\ \ \forall t\in\overline{T}(x^H). \tag{176}$$

Furthermore, $\underline{t}_{j(t,\epsilon;x^H)}(x^H)-\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H)$ is uniformly bounded by z^*, and $j(t,\epsilon;x^H)-(t-z^*)$ by $\overline{q}(x^H)-z^*$, for $t\in\overline{T}(x^H)$; so these two convergences are uniform throughout $\overline{T}(x^H)$. It follows that

$$\lim_{\epsilon\to0}\Big[1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j(t,\epsilon;x^H))+(r-\alpha^L)(\underline{t}_{j(t,\epsilon;x^H)}(x^H)-\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H))-\alpha^L\epsilon}\Big)^{1-\gamma}\Big] \tag{177}$$
$$=\lim_{\epsilon\to0}\Big[1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j(t,\epsilon;x^H))-(r-\alpha^H)\epsilon}\Big)^{1-\gamma}\Big] \tag{178}$$
$$=1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)t}\Big)^{1-\gamma} \tag{179}$$

for all $t\in\overline{T}(x^H)$, and that the convergence of (177) and (178) to (179) is uniform

throughout $\overline{T}(x^H)$.

By (150) and (154),

$$\Lambda^H(x^H)e^{-\alpha^H t} > \Lambda^L(x^H)e^{-\alpha^L t} \;\; \forall t \in [z^*, \overline{q}(x^H)), \tag{180}$$

so $\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)t} > 1$ for all such t. By our assumption of $\gamma > 1$, term (179) is positive for all $t \in \overline{T}(x^H)$.

The net utility gain for H from the shift from y^H to $\tilde{y}^H(\epsilon)$—the sum of (174) across $j \in \epsilon\mathbb{N}$—equals

$$\begin{aligned}
&\frac{1}{\gamma-1}\int_{\overline{T}(x^H)} \left(\Lambda^L(x^H)e^{(r-\alpha^L)\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H)}\right)^{1-\gamma} e^{-\alpha^L(z^*+j(t,\epsilon;x^H))} \\
&\Big[\Big(1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j(t,\epsilon;x^H))+(r-\alpha^L)(\underline{t}_{j(t,\epsilon;x^H)}(x^H)-\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H))-\alpha^L\epsilon}\Big)^{1-\gamma}\Big) \\
&\quad e^{\alpha^L\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H)-\delta^H\underline{t}_{j(t,\epsilon;x^H)}(x^H)} \\
&-\Big(1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)(z^*+j(t,\epsilon;x^H))-(r-\alpha^H)\epsilon}\Big)^{1-\gamma}\Big) \\
&\quad e^{(1-\gamma)(r-\alpha^L)(z^*+j(t,\epsilon;x^H)-\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H))-\delta^H(z^*+j(t,\epsilon;x^H)-\epsilon)+\alpha^L(z^*+j(t,\epsilon;x^H))}\Big]dt.
\end{aligned}$$

From (175) to (180), this converges, as $\epsilon \to 0$, to a value strictly greater than

$$\frac{\Lambda^L(x^H)^{1-\gamma}}{\gamma-1}\int_{\overline{T}(x^H)}\Big(1-\Big(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)}e^{(\alpha^L-\alpha^H)t}\Big)^{1-\gamma}\Big)\Big(e^{(\delta^L-\delta^H)\underline{t}_{t-z^*}(x^H)}-e^{(\delta^L-\delta^H)t}\Big)dt, \tag{181}$$

which is positive. Therefore the total net utility gain is positive; H strictly prefers $\tilde{y}^H(\epsilon)$, for a small ϵ, to y^H, given L's best response. An equilibrium x^H must therefore be at most a measure-zero deviation of one satisfying the polarization condition (158).

We will now find the x^H that, among those satisfying (158) for some $\tilde{t}(x^H)$, uniquely maximizes H's utility given L's best response.

Because L invests his resources until $\tilde{t}(x^H)$ and subsequently allocates them δ^L-optimally, we have

$$x_t^L(x^H) = \begin{cases} 0, & t < \tilde{t}(x^H); \\ \Lambda^L(x^H)e^{(r-\alpha^L)t}, & t \geq \tilde{t}(x^H), \end{cases} \tag{182}$$

where $\Lambda^L(x^H)$ satisfies

$$\int_{\tilde{t}(x^H)}^{\infty} \Lambda^L(x^H)e^{-\alpha^L t}dt = B^L \quad \Longrightarrow \quad \Lambda^L(x^H) = B^L\alpha^L e^{\alpha^L \tilde{t}(x^H)}. \tag{183}$$

Likewise, fixing $\tilde{t}$, denote H's most preferred x^H satisfying (158) with $\tilde{t}(x^H) = \tilde{t}$ by $x^H[\tilde{t}]$. $x^H[\tilde{t}]$ must spend H's resources δ^H-optimally up to $\tilde{t}$. Spending δ^H-optimally up to arbitrarily high $\tilde{t}$ will not be compatible with (154), and thus not with (158), in equilibrium; L will eventually prefer spending to waiting until $\tilde{t}$. Nevertheless, we will find the $\tilde{t}$ that would be optimal for H if H *could* spend δ^H-optimally up to $\tilde{t}$, instead of eventually having to switch to a δ^L-optimal schedule (as in (152)). We will then see that spending δ^H-optimally up to the optimal threshold time, t^*, is indeed compatible with (154) and thus (158) in equilibrium.

So we have

$$x_t^H[\tilde{t}] = \begin{cases} \Lambda^H[\tilde{t}]e^{(r-\alpha^H)t}, & t < \tilde{t}; \\ 0, & t \geq \tilde{t}, \end{cases} \tag{184}$$

where $\Lambda^H[\tilde{t}]$ $(= \Lambda^H(x^H[\tilde{t}]))$ satisfies

$$\int_0^{\tilde{t}} \Lambda^H[\tilde{t}] e^{-\alpha^H t} dt = B^H \quad \Longrightarrow \quad \Lambda^H[\tilde{t}] = \frac{B^H \alpha^H}{1 - e^{-\alpha^H \tilde{t}}}. \tag{185}$$

$(x^H[\tilde{t}], x^L(x^H[\tilde{t}]))$ delivers H payoff

$$\begin{aligned} &\frac{1}{1-\gamma}\Bigg[\int_0^{\tilde{t}} e^{-\delta^H t}\Big(\Big(B^H\alpha^H\big(1-e^{-\alpha^H\tilde{t}}\big)^{-1}e^{(r-\alpha^H)t}\Big)^{1-\gamma}\Big)dt \\ &\qquad + \int_{\tilde{t}}^{\infty} e^{-\delta^H t}\Big(\Big(B^L\alpha^L e^{\alpha^L\tilde{t}}e^{(r-\alpha^L)t}\Big)^{1-\gamma}\Big)\Bigg]dt \\ &= \frac{1}{1-\gamma}\Big[(B^H)^{1-\gamma}(\alpha^H)^{-\gamma}\big(1-e^{-\alpha^H\tilde{t}}\big)^{\gamma} + \frac{(B^L\alpha^L)^{1-\gamma}}{\delta^H+\alpha^L-\delta^L}e^{-\gamma\alpha^H\tilde{t}}\Big]. \end{aligned}$$

From the first order condition in $\tilde{t}$, we find a unique maximum at

$$\tilde{t} = t^* \equiv \ln\Big(\frac{B^H\alpha^H}{B^L\alpha^L}\eta + 1\Big)\big/\alpha^L. \tag{186}$$

Substituting (186) into (185) and (183), we find that the δ^H-optimal allocation rate approaching t^*, and the δ^L-optimal allocation rate at t^*, are respectively

$$\begin{aligned} \Lambda^H[t^*]e^{-\alpha^H t^*} &= B^L\alpha^L/\eta, \\ \Lambda^L(x^H[t^*])e^{-\alpha^L t^*} &= B^L\alpha^L. \end{aligned} \tag{187}$$

Since $\eta < 1$, spending impatient-optimally up to t^* is compatible with (154) in equilibrium, as promised. $x^{H*} \equiv x^H[t^*]$ is thus the unique equilibrium strategy for H among those satisfying (158), and a near-unique equilibrium strategy for H overall.

The proof is completed under "last steps" below.

$\gamma = 1$ **case**

Follow the proof of the $\gamma > 1$ case up to (171). By the reasoning preceding (172), we can decompose the spending shift from $\overline{T}(x^H)$ to $\underline{T}(x^H)$, which constitutes the shift from y^H to $\tilde{y}^H(\epsilon)$, into shifts from $\overline{T}_{j,\epsilon}(x^H)$ to $\underline{T}_{j,\epsilon}(x^H)$ for each $j \in \epsilon\mathbb{N}$; and having done so, H's net utility loss from shift j is bounded above by

$$\begin{aligned}
&\int_{\underline{T}_{j,\epsilon}(x^H)} e^{-\delta^H \underline{t}_{j,\epsilon}(x^H)} \Big(\ln\left(\Lambda^L(x^H) e^{(r-\delta^L)\underline{t}_{j-\epsilon,\epsilon}(x^H)}\right) \\
&\qquad\qquad - \ln\left(\Lambda^H(x^H) e^{(\delta^L-\delta^H)(z^*+j)-\delta^L(\underline{t}_{j,\epsilon}(x^H)+\epsilon)+r\underline{t}_{j,\epsilon}(x^H)}\right)\Big) dt \\
&+ \int_{\overline{T}_{j,\epsilon}(x^H)} e^{-\delta^L(z^*+j-\epsilon)} \Big(\ln\left(\Lambda^H(x^H) e^{(r-\delta^H)(z^*+j-\epsilon)}\right) \\
&\qquad\qquad - \ln\left(\Lambda^L(x^H) e^{(r-\delta^L)(z^*+j)}\right)\Big) dt.
\end{aligned} \tag{188}$$

After rearranging, and by (173), this implies that that H's total net utility loss across all j is bounded above by

$$\begin{aligned}
&\int_{\overline{T}(x^H)} e^{-\delta^L(z^*+j(t,\epsilon;x^H))} \Big[e^{\delta^L \underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H) - \delta^H \underline{t}_{j(t,\epsilon;x^H)}(x^H)} \\
&\Big(\ln\left(\frac{\Lambda^L(x^H)}{\Lambda^H(x^H)}\right) + r\left(\underline{t}_{j(t,\epsilon;x^H)-\epsilon}(x^H) - \underline{t}_{j(t,\epsilon;x^H)}(x^H)\right) + (\delta^H - \delta^L)(z^* + j(t,\epsilon;x^H)) + \delta^L\epsilon\Big) \\
&\quad - e^{(\delta^L-\delta^H)(z^*+j(t,\epsilon;x^H))-\delta^H\epsilon} \Big(\ln\left(\frac{\Lambda^L(x^H)}{\Lambda^H(x^H)}\right) + (\delta^H - \delta^L)(z^* + j(t,\epsilon;x^H)) + (r-\delta^H)\epsilon\Big)\Big] dt.
\end{aligned} \tag{189}$$

By the uniform convergences of (175) and (176), (189) converges to zero as $\epsilon \to 0$. Thus, for any $\ell > 0$, $\exists \bar{\epsilon} > 0$ such that

$$\left| U^H\big((x^H, x^L(x^H))\big) - U^H\big((\tilde{x}^H(\epsilon), x^L(\tilde{x}^H(\epsilon)))\big) \right| < \ell \ \ \forall \epsilon < \bar{\epsilon}. \tag{190}$$

Observe that, as $\epsilon \to 0$, $\tilde{x}^H(\epsilon)$ converges uniformly throughout $[0, \infty)$ to a spend-

ing schedule we might denote $\tilde{x}^H(0)$, which satisfies (150), (154), (168), and (156), and such that, by (190), H is indifferent between $\tilde{x}^H(0)$ and x^H given L's best response.

Consider the class of x^H satisfying (150), (154), (168), and (156). By the reasoning of (182)–(185), the (near-unique) optimal x^H for H in this class, given L's best response—which we may denote x^{H*}—delivers H payoff

$$\begin{aligned}&\int_0^{\tilde{t}} e^{-\delta^H t}\ln\Big(\frac{B^H\delta^H}{1-e^{-\delta^H\tilde{t}}}e^{(r-\delta^H)t}\Big)dt+\int_{\tilde{t}}^{\infty} e^{-\delta^H t}\ln\Big(B^L\delta^L e^{\delta^L\tilde{t}}e^{(r-\delta^L)t}\Big)dt\\ =&\frac{e^{-\delta^H\tilde{t}}}{\delta^H}\Big(\ln(B^L\delta^L)-\frac{\delta^L}{\delta^H}+\delta^H\tilde{t}+1\Big)+\frac{e^{-\delta^H\tilde{t}}-1}{\delta^H}\ln\Big(\frac{1-e^{-\delta^H\tilde{t}}}{B^H\delta^H}\Big)-\frac{1}{\delta^H}+\frac{r}{(\delta^H)^2}\end{aligned}$$

for some $\tilde{t}$. From the first order condition in $\tilde{t}$, we find a unique maximum at

$$\tilde{t}=t^*\equiv\ln\Big(1+\frac{B^H\delta^H}{B^L\delta^L}e^{\frac{\delta^L}{\delta^H}-1}\Big)/\delta^H. \tag{191}$$

Substituting (191) into (185) and substituting both terms into (184), we obtain our expression for x^{H*}.

The proof is completed under "last steps" below.

$\gamma<1$ case

Pick up from the end of the "Preliminaries" section above.

Consider a feasible allocation y^H (and corresponding x^H) satisfying (150) and (154) but not (156). Define

$$\begin{aligned}\hat{\underline{Z}}(z;x^H)&\equiv\int_{\mathcal{T}^H(x^H)\cap[0,z)}e^{-\alpha^L t}dt,\\ \hat{\overline{Z}}(z;x^H)&\equiv\int_{\mathcal{T}^L(x^H)\cap[z,\bar{q}(x^H))}e^{-\alpha^L t}dt.\end{aligned}$$

Since $\hat{\overline{Z}}(z;x^H) - \hat{\underline{Z}}(z;x^H)$ strictly and continuously decreases in z from a positive value at $z = 0$ to a negative value at $z = \overline{q}(x^H)$, there is a unique $z^* \in (0, \overline{q}(x^H))$ such that $\underline{Z}(z^*;x^H) = \overline{Z}(z^*;x^H)$. Let

$$\hat{\underline{T}}(x^H) \equiv \mathcal{T}^H(x^H) \cap [0, z^*),$$
$$\hat{\overline{T}}(x^H) \equiv \mathcal{T}^L(x^H) \cap [z^*, \overline{q}(x^H)).$$

It follows from (150) that $y_t^H \geq \Lambda^L(x^H)e^{-\alpha^L t}\ \forall t \in \mathcal{T}^H(x^H)$.

Choose $\epsilon > 0$. Partition $\hat{\overline{T}}(x^H)$ and define $\hat{\overline{T}}_{j,\epsilon}(x^H)$, $\hat{\underline{t}}_j(x^H)$, $\hat{\underline{T}}_{j,\epsilon}(x^H)$, $\hat{j}(t,\epsilon;x^H)$, and $\hat{S}(x^H)$ analogously to (162)–(166).

By the reasoning up to (168), given any allocation y^H satisfying (150) and (154), there is a corresponding $\tilde{y}^H$ (and $\tilde{x}^H$) also satisfying (150) and (154) but for which we also have

$$\hat{S}(\tilde{x}^H) = \emptyset, \tag{192}$$

such that H is indifferent between y^H and $\tilde{y}^H$ given L's best responses.

Consider a feasible allocation y^H (and corresponding x^H) satisfying (150), (154), and (192) but not (156). Given a choice of interval-length ϵ, consider the allocation $\tilde{y}^H(\epsilon)$ (and corresponding $\tilde{x}^H(\epsilon)$) with

$$\tilde{y}_t^H(\epsilon) = \begin{cases} y_t^H, & t \notin \hat{\overline{T}}(x^H) \cup \hat{\underline{T}}(x^H); \\ 0, & t \in \hat{\underline{T}}(x^H); \\ \Lambda^H(x^H)e^{\alpha^L(\hat{\underline{t}}_{\hat{j}(t,\epsilon;x^H)-\epsilon}(x^H) - z^* - \hat{j}(t,\epsilon;x^H)) - \alpha^H \underline{t}_{\hat{j}(t,\epsilon;x^H)}(x^H)}, & t \in \hat{\overline{T}}(x^H). \end{cases}$$

Note that $\tilde{x}^H(\epsilon)$ satisfies (158) for all ϵ.

To demonstrate that $\tilde{y}^H$ is feasible, let us show that its allocation to each $\hat{\overline{T}}_{j,\epsilon}(x^H)$ is weakly (and indeed strictly) less than y^H's allocation to the corresponding $\hat{\underline{T}}_{j,\epsilon}(x^H)$. From the analogs to (163) and (164), we have the analog to (169). Observe that $t \geq \hat{\underline{t}}_{\hat{j}(t,\epsilon;x^H)-\epsilon}(x^H)\ \forall t \in \hat{\underline{T}}_{j,\epsilon}(x^H)$ and $t < z^* + j\ \forall t \in \hat{\overline{T}}_{j,\epsilon}(x^H)$. Also,

$$y_t^H \geq \Lambda^H(x^H) e^{-\alpha^H \hat{\underline{t}}_j(x^H)}\ \ \forall t \in \hat{\underline{T}}_{j,\epsilon}(x^H).$$

Thus

$$\begin{aligned} \int_{\hat{\overline{T}}_{j,\epsilon}(x^H)} e^{-\alpha^L(z^*+j)}dt &\leq \int_{\hat{\underline{T}}_{j,\epsilon}(x^H)} e^{-\alpha^L \hat{\underline{t}}_j(x^H)}dt \\ \implies \int_{\hat{\overline{T}}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{\alpha^L(\hat{\underline{t}}_{j-\epsilon}(x^H)-z^*-j)-\alpha^H \hat{\underline{t}}_j(x^H)}dt &\leq \int_{\hat{\underline{T}}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{-\alpha^H \hat{\underline{t}}_j(x^H)}dt \\ &< \int_{\hat{\underline{T}}_{j,\epsilon}(x^H)} \Lambda^H(x^H) e^{-\alpha^H t}dt. \end{aligned}$$

Summing across $j \in \epsilon\mathbb{N}$, it follows that, since y^H is feasible, $\tilde{y}^H$ is also feasible.

By calculations precisely analogous to those from (171) to (181)—here simply moving H's spending forward from $\hat{\underline{T}}$ to $\hat{\overline{T}}$, rather than backward from $\overline{T}$ to $\underline{T}$—the total net utility gain for H from the shift from y^H to $\tilde{y}^H(\epsilon)$ converges, as $\epsilon \to 0$, to a value strictly greater than

$$\frac{(\Lambda^L(x^H))^{1-\gamma}}{1-\gamma} \int_{\hat{\overline{T}}(x^H)} \left(\left(\frac{\Lambda^H(x^H)}{\Lambda^L(x^H)} e^{(\alpha^L - \alpha^H)t} \right)^{1-\gamma} - 1 \right) \left(e^{(\delta^L - \delta^H)\hat{\underline{t}}_{t-z^*}(x^H)} - e^{(\delta^L - \delta^H)t} \right) dt.$$

This is positive, like its analog (181): both the coefficient outside the integral and the first factor in the integral have changed sign. Therefore the total net utility gain is positive; H strictly prefers $\tilde{y}^H(\epsilon)$, for sufficiently small ϵ, to y^H given L's best responses.

We have shown that, if $\gamma < 1$, for any feasible allocation y^H with x^H satisfying (150), (154), and (168), but not (156), there is a feasible $\tilde{y}^H$ (and corresponding $\tilde{x}^H$) which H strictly prefers, given L's best responses, with threshold times $\underline{\tilde{t}}(\tilde{x}^H), \overline{\tilde{t}}(\tilde{x}^H)$ such that

$$x_t^L(x^H) = 0 \ \ \forall t \in [\underline{\tilde{t}}(\tilde{x}^H), \overline{\tilde{t}}(\tilde{x}^H)), \quad x_t^H = 0 \ \ \forall t \notin [\underline{\tilde{t}}(\tilde{x}^H), \overline{\tilde{t}}(\tilde{x}^H)). \tag{193}$$

(In the case of the $\tilde{x}^H$ constructed just above, $\underline{\tilde{t}}(\tilde{x}^H) = z^*$ and $\overline{\tilde{t}}(\tilde{x}^H) = \overline{q}(x^H)$.)

Consider a feasible allocation y^H (and corresponding x^H) satisfying (193) for some $\underline{\tilde{t}}(x^H), \overline{\tilde{t}}(x^H)$. Because L, in equilibrium, allocates his resources δ^L-optimally across $[0, \underline{\tilde{t}}(x^H)) \cup [\overline{\tilde{t}}(x^H), \infty)$, we have

$$x_t^L(x^H) = \begin{cases} 0, & t \in [\underline{\tilde{t}}(x^H), \overline{\tilde{t}}(x^H)); \\ \Lambda^L(x^H) e^{(r-\alpha^L)t}, & t \in [0, \underline{\tilde{t}}(x^H)) \cup [\overline{\tilde{t}}(x^H), \infty), \end{cases} \tag{194}$$

up to measure-zero deviations, where $\Lambda^L(x^H)$ satisfies

$$\begin{aligned} B^L &= \int_0^{\underline{\tilde{t}}(x^H)} \Lambda^L(x^H) e^{-\alpha^L t} dt + \int_{\overline{\tilde{t}}(x^H)}^{\infty} \Lambda^L(x^H) e^{-\alpha^L t} dt \\ \implies \quad \Lambda^L(x^H) &= \frac{B^L \alpha^L}{1 + e^{-\alpha^L \overline{\tilde{t}}(x^L)} - e^{-\alpha^L \underline{\tilde{t}}(x^L)}}. \end{aligned} \tag{195}$$

Likewise, fixing $\underline{\tilde{t}}$ and $\overline{\tilde{t}}$, denote H's most preferred x^H satisfying (194) with $\underline{\tilde{t}}(x^H) = \underline{\tilde{t}}$ and $\overline{\tilde{t}}(x^H) = \overline{\tilde{t}}$ by $x^H[\underline{\tilde{t}}, \overline{\tilde{t}}]$. $x^H[\underline{\tilde{t}}, \overline{\tilde{t}}]$ must allocate H's resources δ^H-optimally across $[\underline{\tilde{t}}, \overline{\tilde{t}})$. Spending δ^H-optimally for an arbitrarily long interval will not be compatible with (154), and thus not with (193), in equilibrium; L will eventually prefer spending during the interval. Nevertheless, we will find the $\underline{\tilde{t}}, \overline{\tilde{t}}$ that would be optimal

for H if H *could* spend δ^H-optimally across an arbitrary $[\underline{\tilde{t}}, \overline{\tilde{t}})$, instead of eventually having to switch to a δ^L-optimal schedule (as in (152)). We will then see that the δ^H-optimal schedule satisfying (193) is indeed compatible with (154) and thus (193) in equilibrium.

So we have

$$x_t^H[\underline{\tilde{t}}, \overline{\tilde{t}}] = \begin{cases} \Lambda^H[\underline{\tilde{t}}, \overline{\tilde{t}}]e^{(r-\alpha^H)t}, & t \in [\underline{\tilde{t}}, \overline{\tilde{t}}); \\ 0, & t \in [0, \underline{\tilde{t}}) \cup [\overline{\tilde{t}}, \infty), \end{cases}$$

where $\Lambda^H[\underline{\tilde{t}}, \overline{\tilde{t}}]$ satisfies

$$\int_{\underline{\tilde{t}}}^{\overline{\tilde{t}}} \Lambda^H[\underline{\tilde{t}}, \overline{\tilde{t}}]e^{-\alpha^H t}dt = B^H \quad \Longrightarrow \quad \Lambda^H[\underline{\tilde{t}}, \overline{\tilde{t}}] = \frac{B^H \alpha^H}{e^{-\alpha^H \underline{\tilde{t}}} - e^{-\alpha^H \overline{\tilde{t}}}}. \tag{196}$$

$(x^H[\underline{\tilde{t}}, \overline{\tilde{t}}], x^L(x^H[\underline{\tilde{t}}, \overline{\tilde{t}}]))$ delivers H payoff

$$\left[\int_0^{\underline{\tilde{t}}} e^{-\delta^H t} \Big(\Lambda^L(x^H[\underline{\tilde{t}}, \overline{\tilde{t}}])e^{(r-\alpha^L)t} \Big)^{1-\gamma} dt + \int_{\underline{\tilde{t}}}^{\overline{\tilde{t}}} e^{-\delta^H t} \Big(\Lambda^H[\underline{\tilde{t}}, \overline{\tilde{t}}]e^{(r-\alpha^H)t} \Big)^{1-\gamma} \right.$$
$$\left. + \int_{\overline{\tilde{t}}}^{\infty} e^{-\delta^H t} \Big(\Lambda^L(x^H[\underline{\tilde{t}}, \overline{\tilde{t}}])e^{(r-\alpha^L)t} \Big)^{1-\gamma} dt \right] \frac{1}{1-\gamma}.$$

Rearranging (196) yields

$$\overline{\tilde{t}} = -\ln \Big(e^{-\alpha^H \underline{\tilde{t}}} - \frac{B^H \alpha^H}{\Lambda^H} \Big) / \alpha^H. \tag{197}$$

Substitute (195) for $\Lambda^L(x^H[\underline{\tilde{t}}, \overline{\tilde{t}}])$ and then (197) for $\overline{\tilde{t}}$, so as to reframe H's optimization problem as being over choices of $\underline{\tilde{t}}$ and Λ^H (with $\overline{\tilde{t}}$ endogenous) instead of $\underline{\tilde{t}}$ and

$\bar{\tilde{t}}$ (with Λ^H endogenous). Integrating and simplifying yields

$$\left[\left(\frac{B^L\alpha^L}{1+\left(e^{-\alpha^H\underline{\tilde{t}}}-\frac{B^H\alpha^H}{\Lambda^H}\right)^{\frac{\alpha^L}{\alpha^H}}-e^{-\alpha^L\underline{\tilde{t}}}}\right)^{1-\gamma}\frac{1+\left(e^{-\alpha^H\underline{\tilde{t}}}-\frac{B^H\alpha^H}{\Lambda^H}\right)^{\frac{\alpha^L+\delta^H-\delta^L}{\alpha^H}}-e^{-(\alpha^L+\delta^H-\delta^L)\underline{\tilde{t}}}}{\alpha^L+\delta^H-\delta^L}\right.$$
$$\left.+(\Lambda^H)^{-\gamma}B^H\right]\frac{1}{1-\gamma}. \tag{198}$$

Differentiating with respect to $\underline{\tilde{t}}$ (and then re-introducing Λ^L and $\bar{\tilde{t}}$ in places, for clarity) gives

$$-\frac{1}{\alpha^L+\delta^H-\delta^L}\frac{(\Lambda^L)^{2-\gamma}}{B^L\alpha^L}\left(1+e^{-(\alpha^L+\delta^H-\delta^L)\bar{\tilde{t}}}-e^{-(\alpha^L+\delta^H-\delta^L)\underline{\tilde{t}}}\right)$$
$$+\frac{(\Lambda^L)^{1-\gamma}}{1-\gamma}\left(e^{-(\alpha^L+\delta^H-\delta^L)\underline{\tilde{t}}}-\left(e^{-\alpha^H\underline{\tilde{t}}}-\frac{B^H\alpha^H}{\Lambda^H}\right)^{\frac{\alpha^H-\alpha^L}{\alpha^H}(\gamma-1)}e^{-\alpha^H\underline{\tilde{t}}}\right).$$

The first of these two added expressions is negative. The second is also negative, as can be seen from the fact that it is zero when the B^H term explicitly represented equals zero, and decreases as this term increases.

Thus, for any feasible x^H satisfying (150), (154), (168), and (193) but not

$$\underline{\tilde{t}}(x^H)=0,$$

there is a strictly preferred feasible $\tilde{x}^H$ (with, incidentally, $\Lambda^H(\tilde{x}^H)=\Lambda^H(x^H)$) satisfying all five conditions.

H's favorite strategy in this class is derived for all $\gamma\neq 1$ in (182)–(187). Denote it by x^{H*}, as done there, with $t^*\equiv\bar{\tilde{t}}(x^{H*})$ in the $\gamma<1$ case.

Last steps

Letting $x^{L*} \equiv x^L(x^{H*})$ and

$$M \equiv 1 + \frac{B^H \alpha^H}{B^L \alpha^L}\,\eta,$$

it follows from (182)–(185) and (186) in the $\gamma \neq 1$ cases, and (191) in the $\gamma = 1$ case, that

$$x_t^{H*} = \begin{cases} B^H \alpha^H \frac{M}{M-1} e^{(r-\alpha^H)t}, & t < t^*; \\ 0, & t \geq t^* \end{cases}$$

and

$$x_t^{L*} = \begin{cases} 0, & t < t^*; \\ B^L \alpha^L M^{\frac{\alpha^L}{\alpha^H}} e^{(r-\alpha^L)t}, & t \geq t^* \end{cases}$$

for all $\gamma > 0$, where

$$t^* \equiv \ln(M)/\alpha^H.$$

Given L's strategy $x^L(\cdot)$ (or any measure-zero deviation from it), which is L's unique optimal strategy, H strictly prefers x^{H*} to any alternative spending schedule $x^{H(0)}$ that is a positive-measure deviation from x^{H*} (a "PMD"). This follows from the constructive derivation of x^{H*}, as summarized:

- For any PMD $x^{H(0)}$, we can construct a PMD $x^{H(1)}$ satisfying (150) such that $x^{H(1)} \sim^H x^{H(0)}$.
- For any PMD $x^{H(1)}$ satisfying (150):
  - If $x^{H(1)}$ does not satisfy (151) almost everywhere, we can construct an

$x^{H(1')}$ satisfying (150) and (151) such that $x^{H(1')} \succ^H x^{H(1)}$. $x^{H*} \succ^H x^{H(1')}$, by the full derivation above, so $x^{H*} \succ^H x^{H(1)} \sim^H x^{H(0)}$.

– If $x^{H(1)}$ does satisfy (151) almost everywhere, we can construct a PMD $x^{H(2)}$ satisfying (150) and (151) such that $x^{H(2)} \sim^H x^{H(1)}$.

- For any PMD $x^{H(2)}$ satisfying (150) and (151), we can construct a PMD $X^{H(3)}$ satisfying (150) and (154) such that $x^{H(3)} \sim^H x^{H(2)}$.

- For any PMD $x^{H(3)}$ satisfying (150) and (154), we can construct a PMD $x^{H(4)}$ satisfying (150), (154), and (168) such that $x^{H(4)} \sim^H x^{H(3)}$.

- For any PMD $x^{H(4)}$ satisfying (150), (154), and (168), we can construct an $x^{H(5)}$ satisfying (150), (154), (168), and (158) ($\gamma \geq 1$) or (193) ($\gamma < 1$) such that $x^{H(5)} \gtrsim^H x^{H(4)}$ and $x^{H(5)} \neq x^{H*}$.

- H strictly prefers x^{H*} to all other spending schedules satisfying (150), (154), (168), and (158) ($\gamma \geq 1$) or (193) ($\gamma < 1$). So $x^{H*} \succ^H x^{H(5)} \gtrsim^H x^{H(0)}$.

This completes the result.

A.5 Proof of Proposition 5

Open-loop and Stackelberg schedules

Before beginning, it will be useful to introduce some notation regarding "open-loop" and "Stackelberg" schedules in general, and familiarize ourselves with some of their properties.

Open-loop schedules

Given a nonempty interval $[t, \bar{t})$ and a pair of budgets B^H, B^L to be allocated across it at least one of which is positive, let $x^o[t, \bar{t}, B^L, B^H]$—suppressing some or all of these arguments when clear—denote the (potentially truncated) schedule on which B^H is allocated δ^H-optimally across $[t, t^o)$ and B^L is allocated δ^L-optimally across $[t^o, T)$, where t^o (or $t^o[t, \bar{t}, B^L, B^H]$, etc.) is the point in $[t, \bar{t}]$ that uniquely renders collective spending continuous at t^o. Call this schedule the "open-loop" schedule, and call t^o its "regime-change point".

$$x_s^{H(o)} = \begin{cases} \frac{B^H \alpha^H}{1 - e^{-\alpha^H (t^o - t)}} e^{(r - \alpha^H)(s-t)}, & s \in [t, t^o); \\ 0 & s \in [t^o, \bar{t}), \end{cases} \tag{199}$$

$$x_s^{L(o)} = \begin{cases} 0, & s \in [t, t^o); \\ \frac{B^L \alpha^L}{e^{-\alpha^L (t^o - t)} - e^{-\alpha^L (\bar{t} - t)}} e^{(r - \alpha^L)(s-t)}, & s \in [t^o, \bar{t}). \end{cases}$$

These formulas for x^o, and a proof that it is the unique equilibrium schedule across its interval when the schedule from $\bar{t}$ onward depends only on $\{B_{\bar{t}}^i\}$, follow immediately from (119) and the proof of Proposition 3 (Appendix A.3).

$t^o[B^H, B^L]$ is $\mathcal{C}^1$ in its arguments, by the implicit function theorem and the fact that t^o uniquely satisfies

$$\lim_{s \to t^{o-}} x_s^{H(o)} - x_{t^o}^{L(o)} = 0,$$

whose left-hand side is $\mathcal{C}^1$ in t^o and $\{B^i\}$. Furthermore, since the first term of this difference increases in B^H and decreases in t^o and the second increases in B^L but decreases in t^o, $t^o[B^H, B^L]$ is increasing (decreasing) in B^H (B^L), as is intuitive.

It follows that, for all $s \in [t, \bar{t})$, the open-loop collective spending rate $X_s^o[B^H, B^L]$ strictly increases in B^i for both i. Since from (199) X_s^o is $\mathcal{C}^1$ in t^o and in $\{B^i\}$, and since we found that t^o is itself $\mathcal{C}^1$ in $\{B^i\}$, X_s^o is $\mathcal{C}^1$ in $\{B^i\}$ with respect to its total derivatives (i.e. accounting for the effects of changes in B^i on t^o). So, for each i, $U^i(x^o[B^i, B^{-i}])$ is $\mathcal{C}^1$ in B^i and B^{-i}.

Stackelberg schedules

Given a nonempty interval $[t, T)$ and a pair of budgets B^H, B^L to be allocated across it at least one of which is positive, call a truncated schedule across the interval IC-polarized if there is a "regime-change point" $\tilde{t} \in [t, T]$ such that H is the only spender across $[t, \tilde{t})$ and L allocates B^L δ^L-optimally across $[\tilde{t}, T)$. We will use T rather than $\bar{t}$ to denote the upper end of the interval because we will only consider intervals ending at the end of the game. The "IC" stands for "incentive-compatible", and refers to the fact that if H exhausts her budget at $\tilde{t}$ and L is restricted from spending before $\tilde{t}$, L can have no incentive to do anything other than allocate $B_{\tilde{t}}^L$ δ^L-optimally across $[\tilde{t}, T)$.

Call H's favorite feasible IC-polarized schedule across the interval the "Stackelberg schedule". Denote it by $x^*[t, T, B^H, B^L]$, and its regime-change point by $t^*[t, T, B^H, B^L]$. As with x^o and t^o, we will suppress some or all of these arguments when clear. We will now show that for any t, $T > t$, $B^H > 0$, and $B^L > 0$, $x^*[t, T, B^H, B^L]$ exists and is unique; existence and uniqueness when one budget or both is zero is trivial. We will then note some of its properties.

(The Stackelberg schedule gets its name from the fact that, as we will later see, $x^*[0, \infty, B^H, B^L]$ is the Stackelberg schedule found in Proposition 4. We will not consider whether, on a finite horizon, x^* is the unique equilibrium schedule of a

finite-horizon transformation of the Stackelberg game.)

Let $\tilde{t}$ denote a candidate regime-change point.

By the logic of the proof of Proposition 1, the δ^L-optimal allocation of B_t^L across $[\tilde{t}, T)$ equals

$$x_s^L = \frac{B_t^L \alpha^L}{e^{-\alpha^L(\tilde{t}-t)} - e^{-\alpha^L(T-t)}} e^{(r-\alpha^L)(s-t)}, \quad s \in [\tilde{t}, T). \tag{200}$$

Likewise, the δ^H-optimal allocation of B_t^H across $[t, \tilde{t})$—which, if $t^* = \tilde{t}$, must of course be the allocation adopted under x^*—equals

$$x_t^H = \frac{B_t^H \alpha^H}{1 - e^{-\alpha^H(\tilde{t}-t)}} e^{(r-\alpha^H)(s-t)}, \quad s \in [t, \tilde{t}). \tag{201}$$

Given that L spends nothing until $\tilde{t}$ and then adopts (200), and that H adopts (201), H's favorite location for $\tilde{t}$ is that which maximizes δ^H-discounted utility:

$$\begin{aligned} t^* = \underset{\tilde{t}}{\operatorname{argmax}} \Bigg[& \int_t^{\tilde{t}} e^{-\delta^H(s-t)} u\Big(\frac{B_t^H \alpha^H}{1 - e^{-\alpha^H(\tilde{t}-t)}} e^{(r-\alpha^H)(s-t)} \Big) ds \\ & + \int_{\tilde{t}}^T e^{-\delta^H(s-t)} u\Big(\frac{B_t^L \alpha^L}{e^{-\alpha^L(\tilde{t}-t)} - e^{-\alpha^L(T-t)}} e^{(r-\alpha^L)(s-t)} \Big) ds \Bigg]. \end{aligned} \tag{202}$$

We will not find a closed-form expression for t^*, but we will show that it exists and is unique.

Integrating (202) yields

$$\frac{1}{1-\gamma}\Bigg[\frac{1}{\alpha^H}\Big(1-e^{-\alpha^H(\tilde{t}-t)}\Big)^{\gamma}\big(B_t^H\alpha^H\big)^{1-\gamma} \tag{203}$$
$$+\frac{1}{\delta^H+\alpha^L-\delta^L}e^{-\gamma\alpha^H(\tilde{t}-t)}\Big(1-e^{(\delta^L-\delta^H-\alpha^L)(T-\tilde{t})}\Big)\Big(\frac{B_t^L\alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}\Big)^{1-\gamma}\Bigg],\quad \gamma\neq 1;$$

$$\frac{1}{\delta^H}\Bigg[\bigg(\ln\Big(\frac{B_t^H\delta^H}{1-e^{-\delta^H(\tilde{t}-t)}}\Big)+\frac{r}{\delta^H}-1\bigg)\Big(1-e^{-\delta^H(\tilde{t}-t)}\Big)$$
$$+\bigg(\ln\Big(\frac{B_t^L\alpha^L}{e^{-\delta^L(\tilde{t}-t)}-e^{-\delta^L(T-t)}}\Big)+(r-\delta^L)\Big(\frac{1}{\delta^H}-t\Big)\bigg)\Big(e^{-\delta^H(\tilde{t}-t)}-e^{-\delta^H(T-t)}\Big)$$
$$+(\delta^H-\delta^L)e^{-\delta^H(\tilde{t}-t)}\tilde{t}-(r-\delta^L)e^{-\delta^H(T-t)}T\Bigg],\quad \gamma=1.$$

Differentiating with respect to $\tilde{t}$ then yields

$$\Big(B_t^H\alpha^H\Big)^{1-\gamma}\frac{\gamma}{1-\gamma}\big(1-e^{-\alpha^H(\tilde{t}-t)}\big)^{\gamma-1}e^{-\alpha^H(\tilde{t}-t)} \tag{204}$$
$$+\Big(B_t^L\alpha^L\Big)^{1-\gamma}\Big(1-e^{-\alpha^L(T-\tilde{t})}\Big)^{\gamma-2}e^{-\gamma\alpha^H(\tilde{t}-t)}$$
$$\bigg(\frac{1}{1-\gamma}\Big(e^{-\alpha^L(T-\tilde{t})}-1\Big)+\frac{\alpha^L}{\delta^H+\alpha^L-\delta^L}\Big(1-e^{(\delta^L-\delta^H-\alpha^L)(T-\tilde{t})}\Big)\bigg),\quad \gamma\neq 1;$$

$$e^{-\delta^H(\tilde{t}-t)}\bigg(\ln\Big(\frac{B_t^H\delta^H}{1-e^{-\delta^H(\tilde{t}-t)}}\Big)-\ln\Big(\frac{B_t^L\delta^L}{1-e^{-\delta^L(T-\tilde{t})}}\Big)$$
$$+\frac{\delta^L}{1-e^{-\delta^L(T-\tilde{t})}}\,\frac{1-e^{-\delta^H(T-\tilde{t})}}{\delta^H}\bigg),\qquad \gamma=1,$$

which is continuous in $\tilde{t}$ and approaches a positive value as $\tilde{t}\to t^+$ and a negative value as $\tilde{t}\to T^-$. This guarantees at least one interior solution. Though the second derivative (i.e. the derivative of (204)) is not always negative, it can be proven as follows that a unique $\tilde{t}=t^*$ sets (204) equal to zero.

Rearranging (204) yields

$$Ⓐ_{\tilde{t}}\big(Ⓑ_{\tilde{t}}Ⓒ_{\tilde{t}} + (1 - Ⓑ_{\tilde{t}})Ⓓ_{\tilde{t}} - 1\big), \tag{205}$$

where

$$\begin{aligned}
Ⓐ_{\tilde{t}} &\equiv \frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)} - 1}\left(\frac{B_t^H \alpha^H}{1 - e^{-\alpha^H(\tilde{t}-t)}}\right)^{-\gamma}, \\
Ⓑ_{\tilde{t}} &\equiv \frac{B_t^L \alpha^L \big/ \big(1 - e^{-\alpha^L(T-\tilde{t})}\big)}{B_t^H \alpha^H \big/ \big(e^{\alpha^H(\tilde{t}-t)} - 1\big)}, \\
Ⓒ_{\tilde{t}} &\equiv \big(B_t^L\big)^{-\gamma}\big(\alpha^L\big)^{1-\gamma}\frac{1 - e^{(\delta^L - \alpha^L - \delta^H)(T-\tilde{t})}}{\delta^H + \alpha^L - \delta^L}\Big(1 - e^{-\alpha^L(T-\tilde{t})}\Big)^{\gamma-1} \cdot \left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)} - 1}\right)^{\gamma}, \\
Ⓓ_{\tilde{t}} &\equiv \frac{1}{1-\gamma}\frac{\left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)}-1}\right)^{1-\gamma} - \left(\frac{B_t^L \alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}\right)^{1-\gamma}}{\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)}-1} - \frac{B_t^L \alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}} \cdot \left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)} - 1}\right)^{\gamma}, \quad \gamma \neq 1; \\
&\quad \frac{\ln\left(\frac{B_t^H \delta^H}{e^{\delta^H(\tilde{t}-t)}-1}\right) - \ln\left(\frac{B_t^L \delta^L}{1-e^{-\delta^L(T-\tilde{t})}}\right)}{\frac{B_t^H \delta^H}{e^{\delta^H(\tilde{t}-t)}-1} - \frac{B_t^L \delta^L}{1-e^{-\delta^L(T-\tilde{t})}}} \cdot \frac{B_t^H \delta^H}{e^{\delta^H(\tilde{t}-t)} - 1}, \qquad \gamma = 1.
\end{aligned}$$

$Ⓐ_{\tilde{t}}$ is positive for all $\tilde{t} > t$. To find $\tilde{t}$ that set (205) equal to zero, therefore, we need only focus on the expression within the outer parentheses.

We will first show that (205) is negative if $Ⓑ_{\tilde{t}} \geq 1$. This amounts to the intuitive observation that, if H has set $\tilde{t}$ so high that spending weakly jumps *up* at $\tilde{t}$, because her resources from t to $\tilde{t}$ are spread so thinly and L's resources from $\tilde{t}$ to T are so concentrated, then H does not want to raise $\tilde{t}$ further.

When $Ⓑ_{\tilde{t}} \geq 1$, $Ⓓ_{\tilde{t}}$ (or its limit in the $Ⓑ_{\tilde{t}} = 1$ case) is weakly less than 1. This can be seen from the facts that $u(\cdot)$ is convex, that the first term of $Ⓓ_{\tilde{t}}$ is the slope

of the line segment connecting the points

$$\left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)}-1}, u\left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)}-1}\right)\right), \quad \left(\frac{B_t^L \alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}, u\left(\frac{B_t^L \alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}\right)\right) \tag{206}$$

(or the slope at that point when they are equal), and the second term of $\textcircled{D}_{\tilde{t}}$ divides the first by the derivative at the lower of the two points. When $\textcircled{B}_{\tilde{t}} \geq 1$, (205) is thus weakly less than

$$\textcircled{A}_{\tilde{t}}\,\textcircled{B}_{\tilde{t}}\big(\textcircled{C}_{\tilde{t}} - \textcircled{D}_{\tilde{t}}\big).$$

But

$$\begin{aligned}\textcircled{D}_{\tilde{t}} &> \left(\frac{B_t^L \alpha^L}{1-e^{-\alpha^L(T-\tilde{t})}}\right)^{-\gamma}\left(\frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)}-1}\right)^{\gamma} \\ &= \textcircled{C}_{\tilde{t}}\,\frac{\alpha^L+\delta^H-\delta^L}{\alpha^L}\,\frac{1-e^{-\alpha^L(T-\tilde{t})}}{1-e^{-(\alpha^L+\delta^H-\delta^L)(T-\tilde{t})}} > \textcircled{C}_{\tilde{t}}.\end{aligned}$$

So (205) is negative.

Since $\textcircled{B}_{\tilde{t}}$ is strictly increasing in $\tilde{t}$ from 0 at $\tilde{t}=t$ to ∞ at $\tilde{t}=T$, there is a unique $\bar{\tilde{t}}$ such that $\textcircled{B}_{\bar{\tilde{t}}}=1$, and any $\tilde{t}$ setting (205) equal to 0 must be less than $\bar{\tilde{t}}$. Denote some such $\tilde{t}$ by t^*. We will now show that it is unique in $(t,\bar{\tilde{t}})$.

We will do this by showing that, if $\textcircled{B}_{t^*}\textcircled{C}_{t^*} + (1-\textcircled{B}_{t^*})\textcircled{D}_{t^*} = 1$, this expression is strictly lower for $\tilde{t} \in (t^*, \bar{\tilde{t}})$. This will prove that the $\tilde{t}$ at which this sum equals 1, and thus at which (205) equals 0, is unique.

Fixing $t^* \in (t,\bar{\tilde{t}})$ at which (205)=0, choose $\tilde{t} \in (t^*, \bar{\tilde{t}})$.

$$\begin{aligned}&\textcircled{B}_{t^*}\textcircled{C}_{t^*} + (1-\textcircled{B}_{t^*})\textcircled{D}_{t^*} - \textcircled{B}_{\tilde{t}}\textcircled{C}_{\tilde{t}} - (1-\textcircled{B}_{\tilde{t}})\textcircled{D}_{\tilde{t}} \\ &= \textcircled{B}_{\tilde{t}}\big(\textcircled{C}_{t^*} - \textcircled{C}_{\tilde{t}}\big) + \big(\textcircled{D}_{t^*} - \textcircled{C}_{t^*}\big)\big(\textcircled{B}_{\tilde{t}} - \textcircled{B}_{t^*}\big) + \big(1-\textcircled{B}_{\tilde{t}}\big)\big(\textcircled{D}_{t^*} - \textcircled{D}_{\tilde{t}}\big).\end{aligned} \tag{207}$$

We will show that each of these terms is positive.

$\text{Ⓑ}_{\tilde{t}}$ and $1 - \text{Ⓑ}_{\tilde{t}}$ are positive for all $\tilde{t} \in (t, \bar{\bar{t}})$, by definition of $\bar{\bar{t}}$. Because $\text{Ⓑ}_{\tilde{t}}$ increases in $\tilde{t}$, $\text{Ⓑ}_{\tilde{t}} - \text{Ⓑ}_{t^*}$ is also positive.

Putting aside the positive constant $(B_t^H \alpha^H)^\gamma (B_t^L)^{-\gamma} (\alpha^L)^{1-\gamma} / (\delta^H + \alpha^L - \delta^L)$, $\text{Ⓒ}_{\tilde{t}}$ can be split into three positive terms each of which decreases in $\tilde{t}$:

$$\frac{1 - e^{(\delta^L - \alpha^L - \delta^H)(T - \tilde{t})}}{1 - e^{-\alpha^L (T - \tilde{t})}} \cdot \left(1 - e^{-\alpha^L (T - \tilde{t})}\right)^\gamma \cdot \left(e^{\alpha^H (\tilde{t} - t)} - 1\right)^{-\gamma}.$$

$\text{Ⓒ}_{t^*} - \text{Ⓒ}_{\tilde{t}}$ is thus positive.

To see that $\text{Ⓓ}_{t^*} - \text{Ⓒ}_{t^*}$ is positive, observe first that, for $t^* < \bar{\bar{t}}$, $\text{Ⓓ}_{t^*} > 1$. Its first term is the slope of the line segment connecting points (206) on the graph of $u(\cdot)$ and the second divides it by the derivative of $u(\cdot)$ at what is, given $t^* < \bar{\bar{t}}$ and thus $\text{Ⓑ}_{t^*} < 1$, the higher of the two points. Setting (205)=0 and rearranging, we then have

$$\begin{aligned} \text{Ⓑ}_{t^*} \text{Ⓒ}_{t^*} + (1 - \text{Ⓑ}_{t^*}) \text{Ⓓ}_{t^*} &= 1 \qquad (208) \\ \implies \text{Ⓒ}_{t^*} &= \text{Ⓓ}_{t^*} - \frac{\text{Ⓓ}_{t^*} - 1}{\text{Ⓑ}_{t^*}} \\ &< \text{Ⓓ}_{t^*}. \end{aligned}$$

To see that $\text{Ⓓ}_{t^*} - \text{Ⓓ}_{\tilde{t}}$ is positive, recall once again that $\text{Ⓓ}_{\tilde{t}}$ is the ratio between the slope of the line segment connecting points (206) and the derivative at the higher point. Labeling the points $(p_1, u(p_1))$ and $(p_2, u(p_2))$ respectively, if both p_1 and p_2 were multiplied by some $m < 1$, both the slope of the line segment and the derivative at the higher point would be multiplied by $m^{-\gamma}$, and the ratio would be unchanged.

In the move from $\textcircled{\mathbb{D}}_{t^*}$ to $\textcircled{\mathbb{D}}_{\tilde{t}}$, however, because

$$\frac{\partial}{\partial \tilde{t}} \frac{B_t^H \alpha^H}{e^{\alpha^H(\tilde{t}-t)} - 1} < 0 \text{ but } \frac{\partial}{\partial \tilde{t}} \frac{B_t^L \alpha^L}{1 - e^{-\alpha^L(T-\tilde{t})}} > 0, \tag{209}$$

p_1 is indeed multiplied by some $m < 1$ and the derivative at the higher point multiplied by $m^{-\gamma}$, but p_2 is multiplied by $\overline{m} > 1 > m$. The slope of the line segment is multiplied by less than $m^{-\gamma}$, so the ratio falls.

We have now found that H's favorite feasible IC-polarized schedule x^* is unique. We have also found that $\textcircled{\mathbb{B}}_{t^*} > 1$. That is, if H allocates B_t^H δ^H-optimally across $[t, t^*)$ and L allocates B_t^L δ^L-optimally across $[t^*, T)$, then collective spending weakly falls at t^*. This establishes that $t^* < t^o$.

Precisely as with t^o in the case of the open-loop schedule, we can use the implicit function theorem to establish that $t^*[B^H, B^L]$ is $\mathcal{C}^1$ in both arguments, increasing in B^H and decreasing in B^L. Increases to B^H increase $X_s^*[B^H, B^L]$ for all $s \in [t, T)$, as with $X^o[\cdot]$, though this is no longer obvious: it is proven in the "Final point: Unique equilibrium schedule given placement of first announcement" section below. Now, however, increases to B^L do not necessarily do so.

Overview

First, we will show that the unique equilibrium schedule of the game with finite horizon T resembles the Stackelberg schedule of Proposition 4, x^*, but on a finite horizon. We will also show that this T-horizon schedule converges pointwise almost everywhere to x^* as $T \to \infty$. This will nearly prove part (a), as it will prove that if x^* is an equilibrium schedule, it is the unique limit equilibrium schedule.

Second, we will prove that for any polarized equilibrium σ^* of the infinite-horizon

game, $x(\sigma^*) = x^*$. This will nearly prove part (b), as it will prove that if a polarized equilibrium schedule exists, it is unique and equals x^*.

Finally, we will show that a polarized equilibrium (of the infinite-horizon game) exists. This will conclude the proofs of (a) and (b).

The proof of the unique equilibrium schedule of the game with finite horizon T proceeds by backward induction. Let $\overline{\tau} \equiv \max\{G^{(n)} \cap [0, T)\}$ denote the highest grid point in T-horizon game n. We will therefore begin by finding the unique equilibrium schedule of each subgame beginning at a node $h_{|\overline{\tau}}$.

If $B^i(h_{|\overline{\tau}}) = 0$ for either i, the subgame equilibrium schedule is trivial: any player $-i$ with a positive budget allocates it δ^{-i}-optimally from $\overline{\tau}$ to T. Announcements are irrelevant.

If $B^i(h_{|\overline{\tau}}) > 0$ for both i, the proof of the unique subgame equilibrium schedule following $h_{|\overline{\tau}}$ will itself proceed by backward induction. We will find equilibrium schedules

1. after the second announced node, if any;
2. after the second announcement has been made, if any;
3. after the first announced node (i.e. we will find when the second announcement may occur in equilibrium), if any;
4. after the first announcement has been made, if any; and then
5. from $\overline{\tau}$ before any announcements have been made (i.e. we will find when the first announcement may occur in equilibrium).

Let σ^* be an equilibrium.

Final period

Unique equilibrium schedule after second announced node

The schedule implemented by σ^* following the second announcement after $\overline{\tau}$ follows immediately from (119) and the proof of Proposition 3 (Appendix A.3). Given a type 5 node $h_{|t}$, $x(h_{|t}, \sigma^*)$ allocates $B^H(h_{|t})$ δ^H-optimally across $[t, t^o)$ and $B^L(h_{|t})$ δ^L-optimally across $[t^o, T)$, where the "regime-change point" $t^o \in [t, T]$ uniquely renders collective spending continuous at t^o.

Equilibrium schedules given placement of second announcement

Given a type 4 node $h_{|t}$, let $\xi \equiv \hat{\xi}(h_{|t}) \in (t, T)$ denote the time of the subsequent node, and let $t^o \in [t, T]$ denote the regime-change point of x^o, the open-loop schedule of $\{B_t^i\}$ across $[t, T)$.

If $B^i(h_{|t}) = 0$ for either i, the unique equilibrium schedule across $[t, T)$ is trivial, with any player $-i$ with a positive budget allocating it δ^{-i}-optimally across this interval. We will therefore assume until the end of this subsection of the proof that $B^i(h_{|t}) > 0$ for both i: i.e. that $t^o \in (t, T)$.

We will now show that, for each possible value of ξ, there is an equilibrium schedule following $h_{|t}$. We will then show that the equilibrium schedule we have found is unique.

If $t^o \leq \xi$, x^o is an equilibrium schedule.

If L deviates from implementing x^o, his deviation has no impact on H's spending at any time, because $x^{H(o)}$ spends down B_t^H by ξ. By definition, $x^{L(o)}$ is L's unique best response to $x^{H(o)}$.

If H deviates in a way that maintains $B_\xi^H = 0$, her deviation likewise has no

impact on L's spending at any time, so must lower her payoff, because $x^{H(o)}$ is her unique best response to $x^{L(o)}$. If she deviates such that $B_\xi^H > 0$, she must in equilibrium increase spending across $[\xi, T)$, because she induces the open-loop schedule following ξ with a higher initial value of B^H (positive instead of zero). Because the δ^H-discounted marginal utility to any allocation in this interval (from X^o) is strictly lower than that to an allocation in $[t, t^o)$, and $u(\cdot)$ is convex, any deviation yielding $B_\xi^H > 0$ thus also lowers her payoff.

If $t^o > \xi$, let $\tilde{x}$ denote a schedule that maximizes U^H subject to the conditions that

a) H spends some budget $\leq B_t^H$ δ^H-optimally across $[t, \xi)$, and is the only spender across this interval and

b) the schedule across $[\xi, T)$ is the open-loop schedule across this interval given initial budgets $B_\xi^H, B_\xi^L = B_t^L e^{r(\xi - t)}$, i.e. $x^o[\xi, T, B_\xi^H, B_\xi^L]$.

To prove that such a schedule exists, let $\tilde{x}(B_\xi^H)$ denote the unique $\tilde{x}$ compatible with (a), (b), and a given value of B_ξ^H. Since U^H is continuous in B_ξ^H and the range of feasible values of B_ξ^H is $[0, B_t^H e^{r(\xi - t)}]$, which is compact, there is a U^H-maximizing $\tilde{x}$, by the extreme value theorem. Also, $\tilde{x}$ must satisfy

$$u'(\tilde{x}_t^H) \geq e^{(r - \delta^H)(\xi - t)} \frac{\partial U^H(x^o[\xi, T, B_\xi^H, B_\xi^L])}{\partial B_\xi^H}, \tag{210}$$

with equality if $B_\xi^H > 0$, or else H could increase her payoff from the $\tilde{x}$ baseline, while maintaining (a) and (b), by marginally shifting resources across ξ.

$\tilde{x}$ falls discontinuously at ξ. If it does not and $B_\xi^H = 0$, so that L begins spending immediately, this violates the stipulation that $t^o > \xi$. If it does and $B_\xi^H > 0$, then (210) cannot hold. Given $B_\xi^H > 0$, H must strictly prefer marginal *allocations* at ξ

to marginal allocations just before ξ. This is because, if H reduces spending before ξ and thus increases B_ξ^H, she cannot allocate these resources to periods just after ξ without affecting $t^o[\xi, T, ...]$. Rather, by increasing B_ξ^H, she delays the regime-change point of the open-loop schedule beginning at ξ, increasing spending not only before t^o but also at times after t^o at which the δ^H-discounted marginal utility of resource allocation is lower.

$\tilde{x}$ is an equilibrium schedule. Any deviation at $h_{|t}$ by H that leaves B_ξ^H unchanged would simply consist of a strictly dispreferred allocation of her resources before ξ, and would have no effect on L's schedule. Any deviation by H that changes B_ξ^H, while maintaining δ^H-optimal spending before ξ (and of course the open-loop schedule after ξ), would leave (210) (with equality if $B_\xi^H > 0$) unsatisfied, and so would offer H an opportunity to increase her payoff by shifting resources across ξ in one direction or the other.

Any deviation by L to spending before ξ would decrease spending across $[\xi, T)$. Since the δ^L-discounted marginal utility of allocating resources to any time in this interval (even from x) is higher than that to any time in $[t, \xi)$, this deviation must lower L's payoff.

To see that there are no equilibrium schedules beyond those identified above, first observe that there cannot be t^L, t^H with

$$t \leq t^L < t^H, \quad x_{t^i}^i(h_{|t}, \sigma^*) > 0 \ \ \forall i. \tag{211}$$

To see this, suppose by contradiction that (211) obtains. Since $\alpha^H > \alpha^L$, we must

have either

$$X_{t^H}(h_{|t}, \sigma^*) < e^{(r-\alpha^L)(t^H - t^L)} X_{t^L}(h_{|t}, \sigma^*) \tag{212}$$

$$\text{or} \quad X_{t^H}(h_{|t}, \sigma^*) > e^{(r-\alpha^H)(t^H - t^L)} X_{t^L}(h_{|t}, \sigma^*). \tag{213}$$

If $\xi \leq t^L$, (211) is incompatible with open-loop behavior after the second announcement.

If $\xi > t^H$, marginal reallocations from just after t^L to just after t^H, or vice-versa, do not affect budget sizes at ξ. They therefore do not affect the schedule implemented from ξ onward. But if (212), by the right-continuity of each player's spending, L prefers a marginal reallocation from just after t^L to just after t^H. Likewise, if (213), H prefers a marginal reallocation from just after t^H to just after t^L. Since one of these reallocations must increase a player's payoff across $[t, \xi)$ without affecting it from ξ onward, (211) is incompatible with equilibrium behavior at $h_{|t}$ given $\xi > t^H$.

If $\xi \in (t^L, t^H]$, let

$$\lambda(\epsilon) \equiv \sup_{s \in (\xi - \epsilon, \xi)} X_s, \quad \lambda \equiv \lim_{\epsilon \to 0} \lambda(\epsilon). \tag{214}$$

Because $\lambda(\epsilon)$ is weakly decreasing in ϵ, the limit is defined by the monotone convergence theorem.

If $\lambda < X_\xi$, then H can increase her payoff by increasing spending marginally for some period just before ξ. This marginally reduces B^H_ξ and so marginally reduces spending at all $s \in [\xi, T)$, and spending offers weakly lower δ^H-discounted marginal utility throughout this interval than at ξ.

If $\lambda \geq X_\xi$ and

$$\forall \epsilon > 0 \;\; \exists\, s \in (\xi - \epsilon, \xi) : x_s^L > 0, \tag{215}$$

then L can increase his payoff by reducing spending marginally for some period before ξ. This marginally increases B_ξ^L and so marginally increases spending at all $s \in [\xi, T)$, and allocation offers weakly higher δ^L-discounted marginal utility throughout this interval than at ξ.

If $\lambda \geq X_\xi$ and (215) fails, then either there exists $s \in [t, \xi)$ with $X_s = 0$ or spending across $[t, \xi)$ is not polarized: i.e. there exists $s \in (t^L, \xi)$ with $x_s^H > 0$. Both possibilities are incompatible with equilibrium behavior at $h_{|t}$ because the equilibrium schedule across $[t, \xi)$ must be the open-loop allocation of whatever resources are spent on this interval.

We have now shown that, in equilibrium, there do not exist t^L, t^H satisfying (211). That is, the time t^* at which H exhausts her budget is also the earliest time at which L spends (on $[t, T)$). We also know that the equilibrium schedule is an open-loop schedule before and after ξ. We will now see that these conditions rule out equilibrium schedules other than x^o given $t^o \leq \xi$ and $\tilde{x}$ (a U^H-maximizing schedule satisfying (a) and (b), not shown to be unique) given $t^o > \xi$.

Given $t^o \leq \xi$, a schedule $x \neq x^o$ satisfying the above conditions must feature $t^* < t^o$ or $t^* > t^o$.

If $t^* < t^o$, then $x_s^H > x_s^{H(o)}$ for $s < t^*$ (or else H does not exhaust her budget under x, and could increase her payoff by increasing spending at any time before ξ). Spending at t^* must be continuous, because if it is not, one of the players can increase their payoff by shift spending across t^* without affecting the other's spending at any

time. x^L is then unaffordable for L, since it requires L to begin spending sooner and from a higher present-value spending rate—unless x^L falls discontinuously at ξ, which cannot happen in equilibrium, since L could then increase his payoff by smoothing his spending across $[t^*, T)$.

If $t^* \in (t^o, \xi)$, it can be shown precisely as above that either x^L fails to exhaust L's budget or x^L increases discontinuously at ξ, neither of which is compatible with equilibrium.

If $t^* \geq \xi$ and spending falls discontinuously at ξ, x^L fails to exhaust L's budget. If spending does not fall discontinuously at ξ, then, as shown in the construction of $\tilde{x}$, H can increase his payoff by increasing spending before ξ and lowering B^H_ξ.

Given $t^o > \xi$, if $t^* \geq \xi$ and x does not maximize H's payoff subject to (a) and (b), H can increase her payoff by setting $x^H = \tilde{x}^H$ across $[t, \xi)$ for some schedule $\tilde{x}$ that is U^H-maximizing given (a) and (b).

If $t^* < \xi$, the open-loop schedule is either unaffordable for L even up to ξ or x^L must fall discontinuously at ξ. In the latter case, L could profitably deviate by smoothing his spending across $[t^*, T)$ without affecting H's spending anywhere.

Equilibrium placement of second announcement

Let $h^-_{|t}$ be a type 3 node, and let $x^o \equiv x^o[t, T, B^H_t, B^L_t]$.

If $\xi^L_{\bar{\tau}}(h^-_{|t}) = \emptyset$ (and $\xi^H_{\bar{\tau}}(h^-_{|t}) = t$), it follows from the previous section that L implements x^o for any choice of announcement $\sigma^L(h^-_{|t}) \geq t^o$, and some instantiation of "$\tilde{x}$" for $\sigma^L(h^-_{|t}) < t^o$. Relative to x^o, therefore, any choice of $\sigma^L(h^-_{|t}) < t^o$ induces less spending across $[\sigma^L(h^-_{|t}), T)$ and greater spending across $[t, \sigma^L(h^-_{|t}))$, a reallocation which L disprefers. We must therefore have $\sigma^{L*}(h_{|t}) \geq t^o$, and $x(h^-_{|t}, \sigma^*) = x^o$.

If $\xi^H_{\overline{\tau}}(h^-_{|t}) = \emptyset$ (and $\xi^L_{\overline{\tau}}(h^-_{|t}) = t$), it likewise follows from the previous section that for any choice of $\sigma^H(h^-_{|t})$, H implements an IC-polarized schedule.

Recall that H's favorite feasible IC-polarized schedule x^* is unique, with $t^* < t^o$. We therefore have $\sigma^{L*}_s(h^+_{|t}) = 0$ for $s \in [t, t^*)$, where $h^+_{|t}$ denotes the type 4 node subsequent to $h^-_{|t}$ given $\sigma^H(h^-_{|t}) = t^*$. So H can indeed achieve x^* with this choice of announcement, and cannot achieve a superior schedule.

This identifies the unique equilibrium schedule following $h^-_{|t}$.

Unique equilibrium schedule given placement of first announcement

Let $h_{|\overline{\tau}}$ be a type 2 node and $x^o \equiv x^o[\overline{\tau}, T, B^H_{\overline{\tau}}, B^L_{\overline{\tau}}]$. Note that t^o is interior because we have assumed $B^i_{\overline{\tau}} > 0$ for both i. Let $\xi \equiv \hat{\xi}(h_{|\overline{\tau}})$.

If $\xi^H_{\overline{\tau}}(h_{|\overline{\tau}}) = \xi$ is defined, the unique equilibrium schedule across $[\xi, T)$ is $x^o[\xi, T, B^H_\xi, B^L_\xi]$. So, by a proof identical to that characterizing equilibrium schedules following a type 4 node, any equilibrium schedule following $h_{|\overline{\tau}}$ is IC-polarized.

If $\xi^H_{\overline{\tau}}(h_{|\overline{\tau}}) = \emptyset$, so that $\xi^L_{\overline{\tau}}(h_{|\overline{\tau}}) = \xi$, the unique equilibrium schedule across $[\xi, T)$ is $x^*[\xi, T, B^H_\xi, B^L_\xi]$. Furthermore, because x^* depends only on the initial budgets and not on other features of the history, the unique equilibrium schedule across $[t, \xi)$ is the open-loop schedule for whatever pair of budgets is allocated to this interval. Given this pair of observations, we will now show that any equilibrium schedule following $h_{|\overline{\tau}}$ is IC-polarized.

Suppose $x(h_{|\overline{\tau}}, \sigma^*)$ is not IC-polarized. It must then consist of an open-loop schedule before ξ, on which L spends, followed by a Stackelberg schedule on and after ξ, on which H spends.

If $x(h_{|\overline{\tau}}, \sigma^*)$ is not IC-polarized and $\lambda \geq X_\xi(h_{|\overline{\tau}}, \sigma^*)$, where λ is defined as in (214), we will show that L can increase his payoff by reducing spending before ξ. We will show this in two steps.

First, we will show that the δ^L-discounted utility generated by a resource allocation across the interval before ξ during which L spends is strictly less than the δ^L-discounted utility lost by an equal decrease to $e^{-r(\xi-\overline{\tau})}B_\xi^H$. That is, L would prefer to invest marginal resources until ξ and then to transfer them to H (if this were permitted).

Second, we will show that if L invests to ξ, given that $\sigma^H = \sigma^{H*}$ from ξ onward, he can implement the collective schedule that would result by transferring the resources to H at ξ and then following strategy σ^{L*}. This will imply that the δ^L-discounted utility lost by decreasing H's budget at ξ is weakly less than that lost by equally decreasing L's budget at ξ. In conjunction, these results imply that L strictly prefers to reduce spending before ξ, as desired.

For the first step, given a time $s < \xi$ at which L spends, the δ^L-discounted flow utility to a marginal resource allocation at s equals

$$m^L \equiv e^{(r-\delta^L)(s-t)}(\sigma_s^{L*}(h_{|\overline{\tau}}))^{-\gamma}$$

per unit allocated.

As noted in the discussion of Stackelberg schedules at the beginning of this proof, a decrease to B_ξ^H decreases $t^*[\xi, T, B_\xi^H, B_\xi^L]$.

A decrease to B_ξ^H does not however decrease t^* by enough that

$$x_\xi^H = \frac{B_\xi^H \alpha^H}{1 - e^{-\alpha^H(t^*-\xi)}} \tag{216}$$

weakly rises. To see this, choose two potential values for H's budget at ξ, $\underline{B}^H < \overline{B}^H$, and suppose $t^*[\underline{B}^H] < t^*[\overline{B}^H]$ are such that

$$\frac{\underline{B}^H \alpha^H}{1 - e^{-\alpha^H (t^*[\underline{B}^H] - \xi)}} > \frac{\overline{B}^H \alpha^H}{1 - e^{-\alpha^H (t^*[\overline{B}^H] - \xi)}}. \tag{217}$$

For $Z = A, B, C, D$, let $\overline{\text{Ⓩ}}$ denote the value of Ⓩ at $B^H = \overline{B}^H$, $\tilde{t} = t^*[\overline{B}^H]$, and $t = \xi$ (as ξ is now the beginning of the interval), fixing B^L. Define $\underline{\text{Ⓩ}}$ likewise, but at $B^H = \underline{B}^H$ and $\tilde{t} = t^*[\underline{B}^H]$. If the two sides of (217) were equal, a proof precisely analogous to that from (205) to just following (209) would yield

$$\underline{\text{Ⓑ}}\left(\overline{\text{Ⓒ}} - \underline{\text{Ⓒ}}\right) + \left(\overline{\text{Ⓓ}} - \overline{\text{Ⓒ}}\right)\left(\underline{\text{Ⓑ}} - \overline{\text{Ⓑ}}\right) + \left(1 - \underline{\text{Ⓑ}}\right)\left(\overline{\text{Ⓓ}} - \underline{\text{Ⓓ}}\right) < 0, \tag{218}$$

with $\overline{\text{Ⓒ}} - \underline{\text{Ⓒ}}$, $\underline{\text{Ⓑ}} - \overline{\text{Ⓑ}}$, and $\overline{\text{Ⓓ}} - \underline{\text{Ⓓ}}$ negative, and the other terms positive. Inequality (217) then increases the absolute values of these three differences without affecting the positive terms. So (218) holds.

Thus, when H begins at ξ with budget $\underline{B}^H < \overline{B}^H$, if she exhausts her budget at a time so far below $t^*[\overline{B}^H]$ that (217) holds, then the marginal utility for H to increasing the regime-change point from $t^*[\underline{B}^H]$ is greater than the marginal utility for H to increasing the regime-change point from $t^*[\overline{B}^H]$ given initial budget $\overline{B}^H$. But both must be zero. This concludes the proof that when B^H_ξ is smaller, H's spending path from ξ to the chosen regime-change point $t^*[B^H_\xi]$ is lower.

H's spending path before t^* falls, the point t^* at which spending falls discontinuously itself decreases, and L's spending path after t^* falls. It follows that, as in the open-loop case, a decrease to B^H_ξ decreases the resources allocated to all periods from ξ to T in equilibrium. Furthermore, the δ^L-discounted flow utility to a marginal resource allocation at $s \geq \xi$ increases in s (at proportional rate $\delta^H - \delta^L$) from ξ to

t^*, then jumps and is constant from t^* to T. Since, flow utility is strictly concave in flow spending, the δ^L-discounted utility (as of t) lost by a decrease to $e^{-r(\xi-t)}B^H_\xi$ is strictly more than m^L per unit decreased. This proves that if at t L allocates any resources before ξ, and collective spending weakly falls at ξ, then L would prefer on the margin to invest resources until ξ and then to transfer them to H (if this were permitted).

We have just seen that, when B^H_ξ increases, x^H_ξ increases. For the second step, we will first show that, when B^L_ξ increases, x^H_ξ increases by weakly less than when B^H_ξ increases.

Suppose that it does not. Then, given a partial schedule $x_{|\xi}$, for simplicity let B^i denote $B^i(x_{|\xi})$,

$$U^H(B^H, B^L) \equiv \int_\xi^T e^{-\delta^H(s-\xi)}u(X_t(x_{|\xi}, \sigma^*))ds, \tag{219}$$

and

$$V^H(B^H, B^L, t^*)$$

denote H's payoff following ξ given B^H, B^L, and (not necessarily optimal) regime-change time t^*, as defined by (203) (with ξ, t^* replacing $t, \tilde{t}$). By the envelope theorem, when at ξ H optimally sets the regime-change point at $t^*[B^H, B^L]$, since her payoff after ξ is $\mathcal{C}^1$ in both B^H and the regime-change point, the marginal utility to H of having a larger budget is always the same as what it would be if the regime-

change point were fixed:

$$\frac{\partial U^H}{\partial B^H}(B^H, B^L) = \frac{\partial V^H}{\partial B^H}\big(B^H, B^L, t^*[B^H, B^L]\big) = (x_\xi^H)^{-\gamma}. \tag{220}$$

So, on the assumption that when B_ξ^L increases, x_ξ^H increases by more than when B_ξ^H increases, we have

$$\frac{\partial^2 U^H}{\partial B^{H2}} > \frac{\partial^2 U^H}{\partial B^L \partial B^H}. \tag{221}$$

It can be seen from (202) that if both B^i are multiplied by the same proportion (say m), then $\partial U^H/\partial t^*$ is multiplied by the same proportion for all t^* ($m^{1-\gamma}$). The unique t^* setting $\partial U^H/\partial t^* = 0$ therefore does not change, and spending is multiplied by m at all periods. That is, $t^*[B^H, B^L]$ is h.o.d. 0 and $x^*[B^H, B^L]$ is h.o.d. 1. It follows from this homotheticity that $\partial U^H/\partial B^i$ is in turn multiplied by the same proportion ($m^{-\gamma}$) for both B^i. Formally,

$$\frac{\frac{\partial^2 U^H}{\partial B^L \partial B^H} B^L + \frac{\partial^2 U^H}{\partial B^{H2}} B^H}{\frac{\partial U^H}{\partial B^H}} = \frac{\frac{\partial^2 U^H}{\partial B^{L2}} B^L + \frac{\partial^2 U^H}{\partial B^H \partial B^L} B^H}{\frac{\partial U^H}{\partial B^L}}. \tag{222}$$

By Young's Theorem,

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H} = \frac{\partial^2 U^H}{\partial B^H \partial B^L}.$$

Substituting the left-hand side for the right-hand side into (222) and rearranging,

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H}\left(B^L \frac{\partial U^H/\partial B^L}{\partial U^H/\partial B^H} - B^H\right) + \frac{\partial^2 U^H}{\partial B^{H2}} B^H \frac{\partial U^H/\partial B^L}{\partial U^H/\partial B^H} = \frac{\partial^2 U^H}{\partial B^{L2}} B^L.$$

Then given (221),

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H} \frac{\partial U^H / \partial B^L}{\partial U^H / \partial B^H} + \frac{B^H}{B^L} \frac{\partial^2 U^H}{\partial B^L \partial B^H} \Big(\frac{\partial U^H / \partial B^L}{\partial U^H / \partial B^H} - 1 \Big) < \frac{\partial^2 U^H}{\partial B^{L2}}. \tag{223}$$

Given B^H_ξ, B^L_ξ, if resources are transferred at ξ from L to H, H can always implement the collective schedule that would have obtained without the transfer, by adopting her original schedule until the original regime-change point $t^*[B^H_\xi, B^L_\xi]$ and then following L's original schedule after $t^*[B^H_\xi, B^L_\xi]$ until her funds are exhausted. H can then strictly improve on this schedule by smoothing her spending around $t^*[B^H_\xi, B^L_\xi]$. This implies that

$$\frac{\partial U^H}{\partial B^H} > \frac{\partial U^H}{\partial B^L}. \tag{224}$$

The term of (223) in parentheses is therefore negative. Since flow spending by H at ξ increases as B^L increases, the envelope theorem gives us that

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H} < 0.$$

The second term of the sum in (223) is therefore positive.

In combination with (224), which implies that the coefficient on $\partial^2 U^H / (\partial B^L \partial B^H)$ in the first term of the sum is less than 1, this gives us

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H} < \frac{\partial^2 U^H}{\partial B^{L2}}. \tag{225}$$

By (221) and (225), if marginal resources are transferred from H to L at ξ (from

some initial values of B^H and B^L), $\partial U^H/\partial B^H$ rises by

$$\frac{\partial^2 U^H}{\partial B^L \partial B^H} - \frac{\partial^2 U^H}{\partial B^{H2}} < 0 \tag{226}$$

(i.e. falls) per unit transferred, and $\partial U^H/\partial B^L$ rises by

$$\frac{\partial^2 U^H}{\partial B^{L2}} - \frac{\partial^2 U^H}{\partial B^L \partial B^H} > 0. \tag{227}$$

As seen above with respect to B^H, by the envelope theorem,

$$\frac{\partial U^H}{\partial B^L}(B^H, B^L) = \frac{\partial V^H}{\partial B^L}\big(B^H, B^L, t^*[B^H, B^L]\big).$$

Partially differentiating (203) with respect to B^L, with $\xi, t^*[B^H, B^L]$ replacing $t, \tilde{t}$ respectively, we have

$$\big(B^L\big)^{-\gamma} \frac{1 - e^{(\delta^L - \alpha^L - \delta^H)(T - t^*[B^H, B^L])}}{\delta^H + \alpha^L - \delta^L} e^{-\gamma \alpha^H (t^*[B^H, B^L] - \xi)} \left(\frac{\alpha^L}{1 - e^{-\alpha^L (T - t^*[B^H, B^L])}} \right)^{1-\gamma}.$$

Given (227), there exists an $\epsilon > 0$ such that, when B^H falls by ϵ and B^L increases

by ϵ, this partial derivative rises:

$$\begin{aligned}
&\left(B^L+\epsilon\right)^{-\gamma}\frac{1-e^{(\delta^L-\alpha^L-\delta^H)(T-t^*[B^H-\epsilon,B^L+\epsilon])}}{\delta^H+\alpha^L-\delta^L}\cdot \\
&e^{-\gamma\alpha^H(t^*[B^H+\epsilon,B^L-\epsilon]-\xi)}\left(\frac{\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H+\epsilon,B^L-\epsilon])}}\right)^{1-\gamma} \qquad (228)\\
\geq &\left(B^L\right)^{-\gamma}\frac{1-e^{(\delta^L-\alpha^L-\delta^H)(T-t^*[B^H,B^L])}}{\delta^H+\alpha^L-\delta^L}e^{-\gamma\alpha^H(t^*[B^H,B^L]-\xi)}\left(\frac{\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H,B^L])}}\right)^{1-\gamma}\\
\iff &\frac{(B^L+\epsilon)\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H-\epsilon,B^L+\epsilon])}}\leq\frac{B^L\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H,B^L])}}e^{\alpha^H(t^*[B^H,B^L]-t^*[B^H-\epsilon,B^L+\epsilon])}\times\\
&\left(\frac{\left(1-e^{(\delta^L-\alpha^L-\delta^H)(T-t^*[B^H,B^L])}\right)\Big/\left(1-e^{-\alpha^L(T-t^*[B^H,B^L])}\right)}{\left(1-e^{(\delta^L-\alpha^L-\delta^H)(T-t^*[B^H-\epsilon,B^L+\epsilon]\right])}\right)\Big/\left(1-e^{-\alpha^L(T-t^*[B^H-\epsilon,B^L+\epsilon])}\right)}\right)^{-\frac{1}{\gamma}}.
\end{aligned}$$

Because we have seen that both decreases to B^H and increases to B^L lower t^*,

$$t^*[B^H-\epsilon,B^L+\epsilon]<t^*[B^H,B^L]. \qquad (229)$$

It follows that the fraction in the numerator of the large fraction just above exceeds the fraction in the denominator, and thus that

$$\frac{(B^L+\epsilon)\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H-\epsilon,B^L+\epsilon])}}<e^{\alpha^H(t^*[B^H,B^L]-t^*[B^H-\epsilon,B^L+\epsilon])}\frac{B^L\alpha^L}{1-e^{-\alpha^L(T-t^*[B^H,B^L])}}. \qquad (230)$$

Let Ⓐ_{pre} (for "pre-transfer") denote the value of Ⓐ at B^H, B^L, and $t^*[B^H,B^L]$; Ⓐ_{post} denote the value of Ⓐ at $B^H-\epsilon$, $B^L+\epsilon$, and $t^*[B^H-\epsilon,B^L+\epsilon]$; and $\text{Ⓑ}_{\text{pre/post}}$–$\text{Ⓓ}_{\text{pre/post}}$ likewise. We will now show that, given (226) and (230),

$$\begin{aligned}
&\text{Ⓑ}_{\text{pre}}\left(\text{Ⓒ}_{\text{post}}-\text{Ⓒ}_{\text{pre}}\right)+\left(\text{Ⓓ}_{\text{post}}-\text{Ⓒ}_{\text{post}}\right)\left(\text{Ⓑ}_{\text{pre}}-\text{Ⓑ}_{\text{post}}\right)\\
+&\left(1-\text{Ⓑ}_{\text{pre}}\right)\left(\text{Ⓓ}_{\text{post}}-\text{Ⓓ}_{\text{pre}}\right)>0. \qquad (231)
\end{aligned}$$

We will do this by showing that each of the terms of the sum above is positive.

We know from the discussion surrounding (206) that Ⓑ_{pre} and $1 - \text{Ⓑ}_{\text{pre}}$ are positive. Also, (228)–(229) and (226) give us that $\text{Ⓒ}_{\text{post}} > \text{Ⓒ}_{\text{pre}}$.

It follows from the paragraph ending at (208) that $\text{Ⓓ}_{\text{pre}} > \text{Ⓒ}_{\text{pre}}$. It follows from (230) and (226) that $\text{Ⓑ}_{\text{pre}} > \text{Ⓑ}_{\text{post}}$. It follows in turn from this and the reasoning surrounding (209) that $\text{Ⓓ}_{\text{post}} > \text{Ⓓ}_{\text{pre}}$.

Recall that, by definition, the marginal utility for H of raising t^* from $t^*[B^H, B^L]$ (given that the players' budgets equal B^H, B^L) equals zero. It thus follows from (231), by the calculations preceding (205) and (207), that the marginal utility for H of raising t^* from $t^*(B^H - \epsilon, B^L + \epsilon)$ (given that the players' budgets equal $B^H - \epsilon$ and $B^L + \epsilon$ respectively) is positive. This contradicts the supposition that setting $t^*(B^H - \epsilon, B^L + \epsilon)$ is optimal for H given these budgets.

This completes the proof that (221) is false, i.e. that in fact

$$\frac{\partial^2 U^H}{\partial B^{H2}} \leq \frac{\partial^2 U^H}{\partial B^L \partial B^H},$$

and thus that, when B^L_ξ increases, x^H_ξ increases by weakly less than when B^H_ξ increases.

So, if L reduces spending before ξ, thereby increasing B^L_ξ, he can implement the schedule that would have subsequently have been implemented if the additional resources had been transferred to H at ξ. In combination with the result that L would prefer to reduce spending before ξ if the resources saved would then be transferred to H at ξ, this completes the proof that an equilibrium schedule following $h_{|\overline{\tau}}$ cannot both fail to be IC-polarized and feature a weak decrease at $\xi \equiv \hat{\xi}(h_{|\overline{\tau}})$.

The proof that the schedule implemented by σ^* following $h_{|\overline{\tau}}$ cannot both fail to be IC-polarized and feature $\lambda < X_\xi(h_{|\overline{\tau}}, \sigma^*)$ is more straightforward.

If an equilibrium σ^* implements a non-IC-polarized schedule following $h_{|\overline{\tau}}$, we must have

$$\begin{aligned}\big(X_{\overline{\tau}}(h_{|\overline{\tau}}, \sigma^*)\big)^{-\gamma} &\leq e^{(r-\delta^H)(\xi-\overline{\tau})}\frac{\partial U^H(B^H, B^L)}{\partial B^H} \\ &= e^{(r-\delta^H)(\xi-\overline{\tau})}\big(X_\xi(h_{|\overline{\tau}}, \sigma^*)\big)^{-\gamma},\end{aligned} \tag{232}$$

where, as before, $B^i \equiv B^i_\xi(h_{|\overline{\tau}}, \sigma^*)$ and $U^H(\cdot)$ is defined as in (219). The weak inequality must hold for H not to prefer to increase spending at $\overline{\tau}$ at the cost of lowering her budget at ξ. The equality of the second line follows from the envelope theorem.

But the δ^H-discounted marginal utility of a resource allocation at s, $e^{(r-\delta^H)(s-\overline{\tau})}\big(X_s\big)^{-\gamma}$, weakly falls across $[\overline{\tau}, \xi)$ and, given a jump in collective spending at ξ, falls discontinuously at ξ. So (232) cannot hold.

This completes the proof that any equilibrium schedule following $h_{|\overline{\tau}}$ is IC-polarized.

Unique equilibrium schedule from beginning of period

Let $h^-_{|\overline{\tau}}$ be a type 1 node. Suppose H chooses announcement $\sigma^H(h^-_{|\overline{\tau}}) = t^* \equiv t^*[\overline{\tau}, T, B^H_{\overline{\tau}}, B^L_{\overline{\tau}}]$. We will now show that this subsequently implements the Stackelberg schedule $x^* \equiv x^*[\overline{\tau}, T, B^H_{\overline{\tau}}, B^L_{\overline{\tau}}]$, regardless of $\sigma^L(h^-_{|\overline{\tau}})$ (given that the players will play an equilibrium strategy profile after $h^-_{|\overline{\tau}}$). Since this is H's favorite feasible IC-polarized schedule, this will establish that it is the unique equilibrium schedule following $h^-_{|\overline{\tau}}$. We will let $h^+_{\overline{\tau}}$ denote the node after $h^-_{|\overline{\tau}}$ given the players' announcements.

Regardless of $\sigma^L(h^-_{|\overline{\tau}})$, $\hat{\xi} \equiv \hat{\xi}(h^-_{|\overline{\tau}}) \leq t^*$. Then if σ is an equilibrium, the schedule following $h^+_{|\overline{\tau}}$ must be

- open-loop across $[\overline{\tau}, \hat{\xi})$ (for some non-negative pair of budgets allocated by the players to this interval) and
- IC-polarized.

At $h^+_{|\overline{\tau}}$, therefore, L cannot plan to spend in equilibrium. Because $\hat{\xi} \leq t^* < t^o$, H's optimal open-loop spending plan across $[\overline{\tau}, \hat{\xi})$ that is continuous with L's δ^L-optimal spending up to T would not exhaust H's budget.

Since L spends nothing before $\hat{\xi}$, H can implement $x^*[\overline{\tau}, T, B^H_{\overline{\tau}}, B^L_{\overline{\tau}}]$ by following this schedule up to $\hat{\xi}$ and, at $\hat{\xi}$, if $\hat{\xi} < t^*$, announcing t^* again. This is H's favorite feasible IC-polarized schedule, so this plan must be H's best response to $\sigma^L(h^-_{|\overline{\tau}})$, and what obtains in equilibrium.

Inductive step

Given a grid point $\tau < \overline{\tau}$, suppose that the unique equilibrium schedule across $[\tau', T)$ is Stackelberg. We will show that the unique equilibrium schedule beginning at any pre-announcement node $h_{|\tau}$ is Stackelberg as well.

As before, given a node $h_{|t}$, if $B^i(h_{|t}) = 0$ for some i, the unique subgame equilibrium following $h_{|t}$ is trivial: $-i$ allocates $B^i(h_{|t})$ δ^{-i}-optimally across $[t, T)$. We therefore assume for the rest of this section of the proof that, when studying the game beginning at any node $h_{|t}$, $B^i(h_{|t}) > 0$ for both i.

Unique equilibrium schedule after second announced node

Let $h_{|t}$ be a type 5 node following the second announcement after τ. Let $x^o \equiv x^o[t, T, B^H_t, B^L_t]$ and $x^* \equiv x^*[t, T, B^H_t, B^L_t]$.

A schedule implemented by σ^* following $h_{|t}$ must be IC-polarized, by the proof above of the unique equilibrium schedule following a type 2 node in the last period given $\xi^L > \xi^H$. It must also be Stackelberg across $[\tau', T)$, and, since the Stackelberg schedule following τ' depends only on the budget sizes at τ', it must be open-loop across $[t, \tau')$.

Let $\tilde{t}$ denote the time at which H exhausts her budget on a schedule implemented by σ^*.

If $t^o \leq \tau'$, the unique equilibrium schedule following $h_{|t}$ is x^o.

The fact that no other schedule is an equilibrium schedule follows from polarization. If $\tilde{t} > t^o$, then collective spending increases discontinuously at $\tilde{t}$. Whenever this occurs, L can then increase his payoff by beginning to spend before $\tilde{t}$. If $\tilde{t} < t^o \leq \tau'$, then collective spending must fall discontinuously at $\tilde{t}$, so H can increase her payoff by delaying some spending to after $\tilde{t}$. Neither deviation affects the other player's spending.

Any deviation by L reallocates resources from times with higher to times with weakly lower δ^L-discounted marginal utility to resource allocation without affecting H's spending at any time. A deviation by H that maintains $B^H_{\tau'} = 0$ reallocates resources from times with higher to times with weakly lower δ^H-discounted marginal utility to resource allocation without affecting L's spending at any time. A deviation by H that results in $B^H_{\tau'} > 0$ would increase spending across $[\tau', T)$. Since the δ^H-discounted marginal utility to resource allocation throughout this interval is strictly lower than that to resource allocation before τ', such a deviation must lower H's payoff. x^o is therefore an equilibrium schedule.

If $t^* \geq \tau'$, the unique equilibrium schedule following $h_{|t}$ is x^*.

The fact that no other schedule is an equilibrium schedule follows from polarization. If $\tilde{t} < \tau'$, then collective spending falls discontinuously at $\tilde{t}$, so H can increase her payoff by delaying some spending to after $\tilde{t}$ without affecting L's spending. Given that $\tilde{t} \geq \tau'$, H can implement x^* by and only by setting $x_s^H = x_s^{H*}$ for $s \in [t, \tau')$. Since x^* is H's favorite feasible IC-polarized schedule, H must do this in equilibrium.

Any deviation by H would result in an IC-polarized schedule other than x^*, so H disprefers it. Any deviation by L must take the form of spending before τ'. As shown in the proof above of the unique equilibrium schedule following a type 2 node in the last period given $\xi^L > \xi^H$, L disprefers such a deviation. x^* is therefore an equilibrium schedule.

If $\tau' \in (t^*, t^o)$, the unique equilibrium schedule following $h_{|t}$ consists of H allocating B_t^H δ^H-optimally across $[t, \tau')$ and L allocating B_t^L δ^L-optimally across $[\tau', T)$. Call this the quasi-Stackelberg schedule with regime-change time τ'.

Again, the fact that no other schedule is an equilibrium schedule follows from polarization. If $\tilde{t} < \tau'$, then collective spending falls discontinuously at $\tilde{t}$, so H can increase her payoff by delaying some spending to after $\tilde{t}$ without affecting L's spending. H's favorite schedule among those with $\tilde{t} > \tau'$—given a Stackelberg schedule after τ', and thus a discontinuous drop in collective spending at $\tilde{t}$ and B_t^L spent δ^L-optimally across $[\tilde{t}, T)$—is that in which she spends B_t^H δ^H-optimally across $[t, \tilde{t})$. H's favorite location for $\tilde{t}$, under the restriction that $\tilde{t} \geq \tau'$, then follows from the proof that H has a unique favorite $\tilde{t}$ without this restriction, in the discussion of Stackelberg schedules at the beginning of this proof. There it is shown that, given that the derivative of H's payoff with respect to regime-change time $\tilde{t}$ (there denoted ξ) equals 0 at $\tilde{t} = t^*$, this derivative is negative for all $\tilde{t} > t^*$. Here, since $\tau' > t^*$, it follows that H's favorite location for $\tilde{t}$, under the restriction that $\tilde{t} \geq \tau'$, is τ'. So,

from a strategy profile implementing an IC-polarized schedule in which H does not spend down by τ', H can implement a strictly preferred schedule by spending down δ^H-optimally by τ'.

This implies that H cannot profitably deviate. Since collective spending weakly falls at τ', it is clear that L also cannot profitably deviate. The quasi-Stackelberg schedule with regime-change time τ' is therefore an equilibrium schedule.

Equilibrium schedules given placement of second announcement

Given a type 4 node $h_{|t}$, let $\xi \equiv \hat{\xi}(h_{|t})$ denote the time of the subsequent node. Let $x^o \equiv x^o[t, T, B_t^H, B_t^L]$ and $x^* \equiv x^*[t, T, B_t^H, B_t^L]$. Recall that $t^* < t^o$.

To begin, recall that the unique equilibrium schedule after $h_{|\xi}$ is IC-polarized. Also, since it is determined by $\{B_\xi^i\}$, any equilibrium schedule following $h_{|t}$ is open-loop until ξ.

If $t^o < \xi$, the unique equilibrium schedule following $h_{|t}$ is x^o.

We will first show that no other schedule is an equilibrium schedule.

If L does not spend before ξ, collective spending jumps up at ξ. If H exhausts her budget at ξ, then L can increase his payoff by beginning to spend before ξ. If H spends at ξ, then spending jumps up at ξ, regardless of whether the following schedule is open-loop, Stackelberg, or quasi-Stackelberg. In all three cases, H can increase her payoff by marginally increasing spending to before ξ. In the open-loop case this is because this deviation will shift resources away from all times $[\xi, T)$ to times before ξ, and all the latter offer higher δ^H-discounted marginal utility than all the former; in the Stackelberg case, it follows from the envelope theorem; and in the quasi-Stackelberg case H can reallocate resources from after to before ξ without affecting L's schedule.

Suppose therefore that L spends before ξ. If H spends at ξ, then if collective spending weakly jumps up at ξ, H prefers to increase spending across some interval before ξ during which L spends, by the reasoning just above. If collective spending jumps down at ξ, L wants to decrease his spending marginally before ξ: in the open-loop case because this will constitute a transfer away from all of $[\xi, T)$, and so from lower- to higher- δ^L-discounted marginal utility times; in the Stackelberg case by the proof in the "Final period: Unique equilibrium placement of first announcement" section above; and in the quasi-Stackelberg case because he can then increase reallocate spending to after τ without affecting H's schedule. If instead H exhausts her budget before ξ, then spending must be continuous at ξ, or else L can increase his payoff by smoothing his spending without affecting H's spending. So only x^o can be an equilibrium schedule following $h_{|t}$.

To verify that x^o is in fact an equilibrium schedule, observe that any deviation by H that leaves $B_\xi^H > 0$ transfers resources from before ξ to all times weakly after ξ, which all offer strictly lower δ^H-discounted marginal utility. Any deviation thus lowers her payoff. It is trivial that L has no incentive to deviate.

If $t^* \leq \xi$ and $t^o \geq \xi$, the unique equilibrium schedule following $h_{|t}$ is the quasi-Stackelberg schedule with regime-change time ξ (reducing to the open-loop schedule if $t^o = \xi$ and to the Stackelberg schedule if $t^* = \xi$).

Suppose L spends before ξ. If collective spending jumps down at ξ, L prefers to reallocate marginal spending from before to after ξ regardless of which equilibrium-type prevails from ξ onward, as explained above. If collective spending does not jump down at ξ, then H must spend at ξ (since $t^o > \xi$), and H prefers to reallocate marginal spending from after to before ξ regardless of which equilibrium-type prevails from ξ onward, as explained above.

So in any equilibrium, L does not spend before ξ. H does not spend at or after ξ, because if she does, the equilibrium schedule must be IC-polarized with regime-change time $\tilde{t} > \xi$; H strictly disprefers any such schedule to the quasi-Stackelberg schedule with regime-change time ξ, as explained above; and H can implement the latter by exhausting her budget by ξ.

This is an equilibrium schedule because any deviation by H implements an IC-polarized schedule which she disprefers, and because again L has no incentive to deviate.

If $t^* \in (\xi, \tau')$ and $t^o \leq \tau'$, the unique equilibrium schedule following $h_{|t}$ is H's favorite IC-polarized schedule that is open-loop after ξ: i.e. the schedule denoted $\tilde{x}$ in the proof of the unique equilibrium schedule given the placement of the second announcement in the final period (see the paragraph up to (210)).

Suppose L spends before ξ. Then, as we have seen, because $t^o > \xi$, either spending jumps down at ξ or it does not and H spends at ξ; and in the former case L prefers to reallocate from before to after ξ, and in the latter case H prefers to reallocate in the other direction. So L does not spend before ξ on any equilibrium schedule.

Because $t^o \leq \tau'$, $t^o[\xi, T, B^H_\xi, B^L_\xi] \leq \tau'$ as long as H spends at least as quickly on $[t, \xi)$ as she does on $x^o[t, T, B^H_t, B^L_t]$; so in this case the unique equilibrium schedule following ξ is open-loop. If H spends more slowly on $[t, \xi)$ than $x^{H(o)}[t, T, B^H_t, B^L_t]$, then $B^H_\xi > 0$ and collective spending jumps up at ξ (whichever of the three equilibria obtains), and H prefers to reallocate from after to before ξ as we have seen. So any equilibrium schedule must be IC-polarized and open-loop after ξ.

H's favorite schedule of this form is found (and found to exist and be unique) in the proof of the unique equilibrium schedule given the placement of the second

announcement in the final period. Since it never features a jump up in collective spending at ξ, its regime-change time is before τ'. Given that L spends nothing before ξ, H can unilaterally implement it (as of t) simply by following it with her own spending up to ξ. So no other schedule can obtain in equilibrium.

Any deviation by L to spending before ξ reallocates resources from all times weakly after ξ to some times before. Since the former all offer strictly higher δ^L-discounted marginal utility to resource allocations than the latter, any deviation lowers L's payoff. Any deviation by H that leaves $t^o[\xi, T, B_\xi^H, B_\xi^L] \leq \tau'$ would by construction lower her payoff. Any more extreme deviation would implement a schedule that can be constructed by beginning with a schedule that leaves $t^o[\xi, T, B_\xi^H, B_\xi^L] \leq \tau'$ and reallocating resources from before ξ to some set of times weakly after ξ, all of which (from this baseline) offer strictly lower δ^H-discounted marginal utility. So H has no incentive to deviate either, and the proposed schedule is in fact an equilibrium schedule.

If $t^* \in (\xi, \tau')$ and $t^o > \tau'$, either only the quasi-Stackelberg schedule with regime-change time τ' is the unique equilibrium schedule following $h_{|t}$, only H's favorite IC-polarized schedule that is open-loop following ξ, or both are equilibrium schedules following $h_{|t}$.

L cannot spend before ξ in equilibrium just as in the $t^* \in (\xi, \tau')$, $t^o \leq \tau'$ case. Again, therefore, B_ξ^H determines the equilibrium schedule following ξ. If B_ξ^H is small (including zero), the equilibrium schedule following ξ is open-loop; if larger, it is quasi-Stackelberg with regime-change time τ'; and if larger still, it is Stackelberg, with a regime-change time after τ'.

If H spends so quickly before ξ that the equilibrium schedule following ξ is Stackelberg with a regime-change time after τ', then her (and collective) spending

must jump up at ξ, since $t^* < \tau'$. H then prefers to reallocate from after to before ξ. This leaves the quasi-Stackelberg and open-loop possibilities. H can implement either, as we have seen. In equilibrium, she must implement her favorite of the two, or simply one of the two if they offer her equal payoffs.

Given that she does so, a deviation by L to spending before ξ never increases spending at any time after ξ, so it always shifts resources from times with higher to times with lower δ^L-discounted marginal utility, lowering his payoff. A deviation by H before ξ always lowers her payoff by construction. The proposed equilibrium schedules are therefore in fact equilibrium schedules.

If $t^* \geq \tau'$, x^* is the unique equilibrium schedule following $h_{|t}$.

The proof that L cannot spend before ξ in equilibrium is precisely as in the cases above. Since only H spends before ξ, and any equilibrium schedule after ξ is IC-polarized, any equilibrium schedule is IC-polarized. x^* is H's favorite feasible IC-polarized schedule, and she can implement it by following it up to ξ. So if there is an equilibrium schedule, it must be x^*.

Any deviation by H would implement an alternative, and therefore dispreferred, IC-polarized schedule. Any deviation by L would implement a reallocation from some or all times weakly after ξ to before ξ, which he disprefers. x^* is therefore an equilibrium schedule.

Unique equilibrium schedule incorporating choice of second announcement

Given a type 3 node $h_{|t}$, let $x^o \equiv x^o[t, T, B_t^H, B_t^L]$ and $x^* \equiv x^*[t, T, B_t^H, B_t^L]$.

Suppose $\xi_\tau^L(h_{|t}) = \emptyset$ (and $\xi_\tau^H(h_{|t}) = t$).

If $t^o \leq \tau'$, L implements x^o by choosing announcement $\xi^L \in [t^o, \tau']$; the quasi-

Stackelberg schedule with regime-change time ξ by choosing $\xi^L \in (t^*, t^o)$; and H's favorite schedule that is open-loop after ξ by choosing $\xi^L \in (t, t^*]$ (which is Stackelberg iff $\xi^L = t^*$). L's favorite collective schedule, among those he can implement, is X^o: the others can all be constructed by beginning with X^o and then shifting resources from $[\xi, T)$ to $[t, \xi)$, where the latter all offer weakly higher δ^L-discounted marginal utility than the former. L thus chooses $\xi^L \in [t^o, \tau']$ in equilibrium, and the schedule implemented is x^o.

If $t^* < \tau' < t^o$, L implements the quasi-Stackelberg schedule with regime-change time ξ^L by choosing $\xi^L \in [t^*, \tau']$, and if L chooses $\xi^L \in (t, t^*)$, he implements either (a) the quasi-Stackelberg schedule with regime-change time τ' or (b) H's favorite feasible IC-polarized schedule that is open-loop after ξ^L. If he chooses $\xi^L \in (t, t^*)$ and H weakly prefers (b) to (a), then the regime-change time implemented by (b) cannot be after τ': if it were, then H would prefer to smooth her spending across $[t, \tau')$ rather than letting her spending jump down at ξ, as it must under (b). The collective schedule associated (b) can therefore be constructed by beginning from that associated with (a) and shifting resources from after to before ξ^L. The collective schedule associated with the quasi-Stackelberg schedule with regime-change time $\xi^L < \tau'$ can be constructed likewise. Since such shifts must lower L's payoff, L's favorite implementable schedule is the quasi-Stackelberg schedule with regime-change time τ'. This is the schedule implemented in equilibrium.

If $t^* \geq \tau'$, the equilibrium schedule is x^* regardless of ξ^L.

Suppose $\xi^H_\tau(h_{|t}) = \emptyset$ (and $\xi^L_\tau(h_{|t}) = t$). As shown in the previous subsection, on equilibrium schedules given the placement of the second announcement, the implemented schedule is IC-polarized for any choice of ξ^H. H's favorite feasible IC-polarized schedule is x^*. If $t^* < \tau'$, she can implement it by choosing $\xi^H = t^*$; if $t^* \geq \tau'$, she

implements it with any choice of ξ^H. x^* is therefore always implemented in equilibrium.

Polarization of equilibrium schedules given placement of first announcement

Given a type 2 node $h_{|\tau}$, let $\xi^H \equiv \xi^H_\tau(h_{|\tau})$, and define ξ^L and $\hat{\xi}$ likewise. Let $x^o \equiv x^o[\tau, T, B^H_\tau, B^L_\tau]$ and $x^* \equiv x^*[\tau, T, B^H_\tau, B^L_\tau]$. We will show that any equilibrium schedule following $h_{|\tau}$ is IC-polarized.

If $\hat{\xi} = \xi^H = \xi^L = \tau'$, the equilibrium schedule(s) following $h_{|\tau}$ are those found following a type 5 node $h_{|t}$ (with τ in place of t), all of which are IC-polarized.

If $\hat{\xi} = \xi^H = \xi^L < \tau'$, the equilibrium schedule(s) following $h_{|\tau}$ are those found following a type 4 node $h_{|t}$ (with τ in place of t and $\hat{\xi}$ in place of $\hat{\xi}(h_{|t})$), all of which are IC-polarized.

If $\hat{\xi} = \xi^L < \xi^H$, the equilibrium schedule following $\hat{\xi}$ is Stackelberg. As we have seen, this implies that the equilibrium schedule following $h_{|\tau}$ is open-loop if $t^o \leq \hat{\xi}$, quasi-Stackelberg with regime-change time $\hat{\xi}$ if $t^* < \hat{\xi} < t^o$, and Stackelberg if $\hat{\xi} \leq t^*$. In every case, it is IC-polarized.

If $\hat{\xi} = \xi^H < \xi^L$, the equilibrium schedule following $\hat{\xi}$ is open-loop if $t^o[\hat{\xi}, T, B^H_{\hat{\xi}}, B^L_{\hat{\xi}}] \leq \tau'$, quasi-Stackelberg with regime-change time τ' if $t^*[\hat{\xi}, T, B^H_{\hat{\xi}}, B^L_{\hat{\xi}}] < \tau' < t^o[\hat{\xi}, T, B^H_{\hat{\xi}}, B^L_{\hat{\xi}}]$, and Stackelberg if $t^*[\hat{\xi}, T, B^H_{\hat{\xi}}, B^L_{\hat{\xi}}] \geq \tau'$. This is precisely analogous to the equilibrium schedules following a type 5 node $h_{|t}$ (with $\hat{\xi}$ in place of t). The equilibrium schedule(s) following $h_{|\tau}$ are those found following a type 4 node $h_{|t}$ (with τ in place of t and $\hat{\xi}$ in place of $\hat{\xi}(h_{|t})$), all of which are IC-polarized.

Unique equilibrium schedule from beginning of period

Given a type 1 node $h^-_{|\tau}$, let $x^o \equiv x^o[\tau, T, B^H_\tau, B^L_\tau]$ and $x^* \equiv x^*[\tau, T, B^H_\tau, B^L_\tau]$, let ξ^i denote $\sigma^{i*}(h^-_{|t})$, and let $h^+_{|\tau}$ denote the subsequent (type 2) node.

If $t^* \geq \tau'$, suppose $\xi^H = \tau'$. If $\xi^L = \tau'$, the unique equilibrium schedule following $h^+_{|\tau}$ is Stackelberg, as following a type 5 node with $t^* > \tau'$. If $\xi^L < \tau'$, the unique equilibrium schedule following ξ^L is Stackelberg; so, since $t^* > \xi^L$, the unique equilibrium schedule following $h^+_{|\tau}$ is also Stackelberg.

If $t^* < \tau'$, suppose $\xi^H = t^*$. If $\xi^L \geq t^*$, the unique equilibrium schedule following ξ^H is open-loop, quasi-Stackelberg, or Stackelberg, in the usual way; L does not spend before ξ^H in equilibrium, because $t^o > \xi^H$; and H can thus implement x^* by spending her budget δ^H-optimally across $[\tau, t^*)$. If $\xi^L < t^*$, the unique equilibrium schedule following ξ^L is Stackelberg; so, since $t^* > \xi^L$, the unique equilibrium schedule following $h^+_{|\tau}$ is also Stackelberg.

Since in equilibrium the schedule following $h^+_{|\tau}$ must be polarized, and since at $h^-_{|\tau}$ H can implement her favorite feasible IC-polarized schedule x^* with the appropriate strategy, x^* is the unique equilibrium schedule following $h^-_{|\tau}$.

A limit equilibrium schedule must be Stackelberg

For the rest of the proof, x^* and t^* will be as defined by Proposition 4.

Continuous time

Having completed the inductive step, we have proven that, fixing horizon T, the equilibrium schedule of game n is unique and Stackelberg. Denote this schedule by $x^{[T]}$. $x^{[T]}$ is independent of n, so it is also the unique equilibrium schedule in continuous time.

$x^{[T]}$ converges pointwise almost everywhere to x^* as $T \to \infty$

Let $U^H[\tilde{t}, T]$ denote H's payoff from a quasi-Stackelberg schedule with regime-change time $\tilde{t}$ across $[0, T)$. Let $f(\tilde{t}, T)$ denote the derivative of U^H with respect to $\tilde{t}$, as given by (204) with 0 replacing t, expressed as a function of $\tilde{t}$ and T. Observe that it is defined and continuous in both variables over $T \in (0, \infty]$ and $\tilde{t} \in (0, T)$.

Fixing B^H and B^L, let $t^*(T)$ denote the value of $\tilde{t}$ that sets (204) equal to 0 given horizon T. Recall that $t^*(T)$ is unique for all $T > 0$, including $T = \infty$. It can be found analytically that $t^*(\infty) = t^*$.

Fixing $\tilde{t}$, when T rises, $\text{Ⓑ}_{\tilde{t}}$ falls and $\text{Ⓒ}_{\tilde{t}}$ and $\text{Ⓓ}_{\tilde{t}}$ rise. By (207), as we have seen, this implies that $f(\tilde{t}, T)$ rises. It follows that $t^*(T)$ increases in T.

By the monotone convergence theorem, $\overline{t}^* \equiv \lim_{T\to\infty} t^*(T)$ is defined. Furthermore, because $f(\tilde{t}, T)$ is continuous, $f(\overline{t}^*, \infty) = \lim_{T\to\infty} f(t^*(T), T)$. This limit equals zero, since each of its elements equals zero by definition of $t^*(\cdot)$. Therefore $\overline{t}^* = t^*$.

It is then easy to verify that $x_t^{[T]}$ converges pointwise to x_t^* for all $t \neq t^*$.

A polarized equilibrium schedule must be Stackelberg

Let σ^* be a polarized equilibrium.

Let $h_{|t}$ be a post-announcement node. Suppose $\sigma_s^{L*}(h_{|t}) > 0$ for some s, and let $\underline{s} \equiv \inf\{s : \sigma_s^{L*}(h_{|t}) > 0\}$. (Recall that $\sigma^{L*}(h_{|t})$ is a spending plan for L across $[t, \hat{\xi}(h_{|t}))$, so $\underline{s} \in [t, \hat{\xi}(h_{|t})]$.) By definition of polarized equilibrium, $x_s^H(h_{|t}, \sigma^*) = 0$ for all $s \geq \underline{s}$ (the $s = \underline{s}$ case following from the assumption that spending plans are right-continuous). For σ^{L*} to be an optimal strategy for L, therefore, we must have $x_s^L(h_{|t}, \sigma^*) = B_{\underline{s}}^L e^{(r-\alpha^L)(s-\underline{s})}$ for all $s \geq \underline{s}$. Since σ^* is a polarized equilibrium and L begins spending before the next node, if $B^H(h_{|t}) > 0$, σ^{H*} must allocate $B^H(h_{|t})$

δ^H-optimally across $[t, \underline{s})$; and collective spending cannot be discontinuous at $\underline{s}$, or else restricting spending to before (weakly after) $\underline{s}$ cannot be a best response for H (L). So, if $B^H(h_{|t}) > 0$, we can only have $\sigma^{L*}(h_{|t}) \neq 0$ if the schedule implemented by σ^* following $h_{|t}$ is open-loop with a regime-change time before $\hat{\xi}(h_{|t})$. This in turn is possible only if $t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})] < \hat{\xi}(h_{|t})$:

$$\begin{aligned} & t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})] \geq \hat{\xi}(h_{|t}) \\ \implies & \sigma^{L*}(h_{|t}) = 0 \qquad \text{for all post-announcement nodes } h_{|t}. \end{aligned} \tag{233}$$

Suppose that σ^H is such that, at any pre-announcement node $h_{|t}$ with $\xi^H_{\tau(t)}(h_{|t}) = \emptyset$ and $B^H(h_{|t}) > 0$, H chooses announcement

$$\sigma^H(h_{|t}) = \min(t^*[t, \infty, B^H(h_{|t}), B^L(h_{|t})], \tau'(t));$$

and that any post-announcement node $h_{|t}$ with $t^*[t, \infty, B^H(h_{|t}), B^L(h_{|t})] \geq \hat{\xi}(h_{|t})$, H chooses spending plan

$$\sigma^H(h_{|t}) = x^{H*}_{[t, \hat{\xi}(h_{|t}))}[t, \infty, B^H(h_{|t}), B^L(h_{|t})].$$

For any σ^{L*} satisfying (233), $x(\sigma^H, \sigma^{L*}) = x^*$. This can be seen by observing that (σ^H, σ^{L*}) implements x^* following all post-announcement nodes:

- At any type 2 node $h_{|t}$ with $B^H(h_{|t}) > 0$,

$$\hat{\xi}(h_{|t}) \leq t^*[t, \infty, B^H(h_{|t}), B^L(h_{|t})] < t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})],$$

  so in equilibrium H is the only spender across $[t, \hat{\xi}(h_{|t}))$, and follows the Stackelberg schedule across this interval.

- H exhausts her budget at $\hat{\xi}(h_{|t})$ only if $\hat{\xi}(h_{|t}) = t^*[t, \infty, B^H(h_{|t}), B^L(h_{|t})]$. So at a type 4 node $h_{|t}$, either $B^H(h_{|t}) = 0$ and the Stackelberg schedule has been followed across $[\tau(t), t)$ (and will be implemented by L across $[t, \infty)$), or $B^H(h_{|t}) > 0$ and $\xi^H_{\tau(t)}(h_{|t}) = \emptyset$, in which case the next node will occur weakly before the Stackelberg regime-change time, and H will be the only spender until the next node, following the Stackelberg schedule.

- A type 5 node $h_{|t}$ can be reached only at the Stackelberg regime-change time, in which case $B^H(h_{|t}) = 0$, the Stackelberg schedule has been followed across $[\tau(t), t)$, and it will be implemented by L across $[t, \infty)$.

Since H can implement x^* when L employs any strategy σ^{L*} satisfying (233), since x^* is her favorite feasible IC-polarized schedule, and since $x(\sigma^*)$ is a feasible IC-polarized schedule for any polarized equilibrium σ^*, there is no polarized equilibrium σ^* with $x(\sigma^*) \neq x^*$.

A polarized equilibrium exists

Let σ^{H*} satisfy the conditions imposed on σ^H in the previous subsection, and define it fully by stipulating that, for a pre-announcement node $h_{|t}$ with $\xi^H_{\tau(t)}(h_{|t}) = \emptyset$ and $B^H(h_{|t}) = 0$, $\sigma^{H*}(h_{|t}) = \tau'(t)$, and for a post-announcement node $h_{|t}$ with $t^*[t, \infty, B^H(h_{|t}), B^L(h_{|t})] < \hat{\xi}(h_{|t})$, H allocates $B^H(h_{|t})$ δ^H-optimally across $[t, t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})])$ if $t^o \leq \tau'(t)$ and across $[t, \tau'(t)]$ otherwise.

Let σ^{L*} satisfy (233); for post-announcement nodes $h_{|t}$ with $t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})] < \hat{\xi}(h_{|t})$,

$$\sigma^{L*}(h_{|t}) = x^{L(o)}_{[t, \hat{\xi}(h_{|t}))}[t, \infty, B^H(h_{|t}), B^L(h_{|t})];$$

and for pre-announcement nodes $h_{|t}$ with $\xi^L_{\tau(t)}(h_{|t}) = \emptyset$, $\sigma^{L*}(h_{|t}) = \tau'(t)$.

Because σ^{L*} satisfies (233), and for any post-announcement node $h_{|t}$ with $t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})] < \tau'(t)$ we have $\sigma^{H*}_s(h_{|t}) = 0$ for all $s \geq t^o[t, \infty, B^H(h_{|t}), B^L(h_{|t})]$, $x_{[t,\infty)}(h_{|t}, \sigma^*)$ is polarized for all nodes $h_{|t}$. Therefore, if σ^* is an equilibrium, it is a polarized equilibrium.

Given that L adopts strategy σ^{L*}, $x((\sigma^H, \sigma^{L*}))$ is IC-polarized for any σ^H. Since $x(\sigma^*) = x^*$, which is H's favorite feasible IC-polarized schedule, σ^{H*} is a best response for H to σ^{L*}.

Following the node-cases of the "inductive step" above, it can be verified that for any grid point τ, if the players will play a strategy profile that implements the Stackelberg schedule from τ' onward (as σ^* does), and if H plays σ^{H*} across $[\tau, \tau')$, σ^{L*} is a best response for L for every possible node that might arise across $[\tau, \tau')$. The proofs of L's best responses through the inductive step rely nowhere on the assumption that T is finite.

When $\gamma < 1$, infinite-horizon game n is continuous at infinity (for any n), since the range of feasible payoffs at any period is bounded above and below and the stream of flow payoffs cannot grow more quickly than the players' discount rates, by (111). By the one-shot deviation principle, therefore, σ^{L*} is a best response for L to σ^{H*}. This establishes that σ^* is an equilibrium of the infinite-horizon game when $\gamma < 1$.

When $\gamma \geq 1$, however, the game is not continuous at infinity. We must verify explicitly that, given that H plays σ^{H*}, at any node $h_{|t}$ L has no incentive to deviate permanently from σ^{L*} to an alternative strategy σ^L.

σ^{L*} maximizes L's forward-looking optimization problem at all nodes $h_{|t}$ with $B^H(h_{|t}) = 0$. Deviation to σ^L therefore can only be profitable for L at a node $h_{|t}$

with $B^H(h_{|t}) > 0$. Suppose by contradiction that there is such an $h_{|t}$ and σ^L.

If $\sigma \equiv (\sigma^{H*}, \sigma^L)$ is such that

$$\exists \tau \in G^{(n)} : B^H(h) = 0, \quad h \equiv h^-_{|\tau}(h_{|t}, \sigma), \tag{234}$$

then σ^L offers L a lower payoff than σ^{L*} at $h_{|t}$. This can be seen by backward induction. First, because of the unique optimality for L of the schedule implemented by σ^{L*} from h onward, a permanent deviation to σ^L cannot offer L higher utility than a deviation to σ^L until h followed by a reversion to σ^{L*}, and must offer L lower utility if $x(h, \sigma) \neq x(h, \sigma^*)$. Then, given that the players play σ^* from τ onward, and given that H plays σ^{H*} from $\underline{\tau}$ onward (where $\underline{\tau}$ denotes the grid point before τ), it follows from the "inductive step" of the proof of the unique limit equilibrium that any strategy $\tilde{\sigma}^L$ among L's most preferred strategies at any node $h_{|s}$ with $s \in [\underline{\tau}, \tau)$ must satisfy $x(h_{|s}, \tilde{\sigma}) = x(h_{|s}, \sigma^*)$. By induction, deviation to a strategy σ^L satisfying (234) is weakly undesirable for L at any node.

Consider a strategy $\tilde{\sigma}^L$ not satisfying (234). We will assume by contradiction that permanently deviating to $\tilde{\sigma}^L$ at $h_{|t}$ increases L's payoff, and show that this implies that L's payoff to playing $\tilde{\sigma}^L$ from time τ onward is $-\infty$ for some grid point $\tau > t$.

By the reasoning above, for any grid point $\tau > t$, L's payoff to playing $\tilde{\sigma}^L$ at all nodes $h_{|s}$ with $s \in [t, \tau)$, and σ^{L*} from τ onward, is weakly less than his payoff to playing σ^{L*} from $h_{|t}$ onward. So, in general denoting L's continuation payoff to playing strategy σ^L from grid point $\tau > t$ onward after playing $\tilde{\sigma}^L$ at all nodes $h_{|s}$ with $s \in [t, \tau)$ by

$$C(\sigma^L, \tau) \equiv \int_\tau^\infty e^{-\delta^L(s-\tau))} u\big(X_s\big(h^-_{|\tau}(h_{|t}, \tilde{\sigma}), (\sigma^{H*}, \sigma^L)\big)\big) ds,$$

and denoting L's payoff to playing σ^L from $h_{|t}$ onward by $C(\sigma^L, t)$, we have

$$C(\sigma^{L*}, t) \geq C(\tilde{\sigma}^L, t) + e^{-\delta^L(\tau - t)}\big(C(\sigma^{L*}, \tau) - C(\tilde{\sigma}^L, \tau)\big)$$
$$\implies C(\tilde{\sigma}^L, \tau) - C(\sigma^{L*}, \tau) \geq e^{\delta^L(\tau - t)}\big(C(\tilde{\sigma}^L, t) - C(\sigma^{L*}, t)\big). \tag{235}$$

If $\tilde{\sigma}^L$ is a profitable deviation for L at $h_{|t}$, the right-hand side of (235) is positive, so the difference in continuation payoffs as a function of $\tau - t$ must be "fast-growing", which we will define to mean asymptotically bounded below by $c_0 e^{\delta^L(\tau-t)}$ for some constant $c_0 > 0$. $C(\tilde{\sigma}^L, \tau)$ can never exceed the continuation payoff for L at τ obtained if both parties invest all funds from t to τ and subsequently disburse them δ^L-optimally. This continuation payoff plateaus if $\gamma > 1$, and grows linearly in $\tau - t$ at absolute rate r if $\gamma = 1$. (See the payoff expression from Proposition 1, substituting $B(h_{|t})e^{r(\tau-t)}$ for B.) For the difference in continuation payoffs to be fast-growing, therefore, $C(\sigma^{L*}, \tau)$ must eventually be negative and its absolute value fast-growing.

Since σ^* implements the Stackelberg schedule from any pre-announcement grid node onward, $C(\sigma^{L*}, \tau) = U^L(x^*[\tau])$, where $x^*[\tau] \equiv x^*[\tau, \infty, B^H(h^-_{|\tau}(h_{|t}, \tilde{\sigma})), B^L(h^-_{|\tau}(h_{|t}, \tilde{\sigma}))]$. An expression for this payoff can be found analytically by integrating

$$\int_\tau^\infty e^{-\delta^L(s-\tau)} u(x^*_s[\tau])ds,$$

because the regime-change point $t^*[\tau]$ can be found analytically in this setting where the horizon is infinite (see Proposition 4). The result can be written in terms of the

collective initial budget B_τ and L's initial budget share b_τ as

$$\frac{B_\tau^{1-\gamma}}{1-\gamma} \cdot \frac{\left(\alpha^L\right)^{-\gamma}}{\alpha^H\alpha^L\gamma + (\delta^H-\delta^L)^2(1-\gamma)} \cdot \Big(\alpha^H\alpha^L\gamma\big(b_\tau + (1-b_\tau)\frac{\alpha^H}{\alpha^L}\eta\big)^{1-\gamma} + (\delta^H-\delta^L)^2 \\ (1-\gamma)b_\tau^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}}\big((b_\tau + (1-b_\tau)\frac{\alpha^H}{\alpha^L}\eta\big)^{-\frac{\alpha^L}{\alpha^H}\gamma}\Big), \quad \gamma > 1; \tag{236}$$

$$\frac{1}{\delta^L}\Big(\frac{(\delta^H-\delta^L)^2}{\delta^H\delta^L}\Big(1 + \frac{(1-b_\tau)\delta^H}{b_\tau\delta^L}\eta\Big)^{-\frac{\delta^L}{\delta^H}} + \ln(B_\tau) + \ln\left(b_\tau\delta^L + (1-b_\tau)\delta^H\eta\right) \\ + \frac{\delta^H-\delta^L}{\delta^H} + \frac{r-\delta^H}{\delta^L}\Big), \qquad \gamma = 1.$$

If $\gamma > 1$, the coefficient on $B_\tau^{1-\gamma}$ is negative and is bounded below across $b_\tau \in [0,1]$. For $C(\sigma^{L*},\tau)$ to be fast-growing, therefore, $B_\tau^{1-\gamma}$ must be fast-growing. B_τ must thus eventually be bounded above by $c_1 e^{\frac{\delta^L}{1-\gamma}(\tau-t)}$ for some $c_1 > 0$. Because the spending rate cannot sustainably shrink more slowly than the collective budget, the discounted continuation payoff to adopting σ^L from a sufficiently large grid point $\tau > t$ must likewise be bounded above, for some $c_2 > 0$, by

$$\int_\tau^\infty e^{-\delta^L(s-t)}\frac{\left(c_2 e^{\frac{\delta^L}{1-\gamma}s}\right)^{1-\gamma}}{1-\gamma}ds = -\infty.$$

If $\gamma = 1$, the terms added to $\ln(B_\tau)$ are likewise bounded across $b_\tau \in [0,1]$. For $C(\sigma^{L*},\tau)$ to be fast-growing, therefore, $\ln(B_\tau)$ must be (again, negative and) fast-growing, so B_τ must eventually be bounded above by a function $f(\tau)$ that falls superexponentially to zero quickly enough that $\ln(f(\tau))$ is (negative and) fast-growing. Again, because the spending rate cannot sustainably shrink more slowly than the collective budget, the discounted continuation payoff to following σ^L from sufficiently

large τ must be bounded above, for some $c_1, c_2 > 0$, by

$$\int_{\tau}^{\infty} e^{-\delta^L(s-t)} \ln\left(c_1 f(s)^{c_2}\right) ds = -\infty.$$

This contradicts the assumption that σ^L is a profitable deviation at $h_{|t}$ from σ^{L*}, whose payoff is well-defined and finite (as confirmed by (236), recalling conditions (111) and (120)). σ^* is thus an equilibrium of infinite-horizon game n for all n.

A.6 Proof of Proposition 6

Preliminaries

Following the notation of Appendix A.5, we will let $x^*[t, T, B^H, B^L]$ denote the Stackelberg schedule beginning at t with budgets B^H and B^L and ending at T, dropping some or all of the arguments where clear. X^*, etc. are defined likewise. We will also let $b^H \equiv B^H/B = 1 - b$.

We begin by showing that the frontier of efficient payoffs is strictly concave.

Let $U^0 = (U^{H(0)}, U^{L(0)})$ and $U^1 = (U^{H(1)}, U^{L(1)})$ be two efficient payoff profiles, and let $X^{(0)}$ and $X^{(1)}$ be collective schedules which attain these payoff profiles.

The mixture collective schedule $X^{(\zeta)}$, defined by $X_t^{(\zeta)} = \zeta X_t^{(1)} + (1-\zeta)X_t^{(0)}$, is feasible:

$$\begin{aligned}
&\int_0^{\infty} e^{-rt}\left(\zeta X_t^{(1)} + (1-\zeta)X_t^{(0)}\right) dt \\
&= \zeta \int_0^{\infty} e^{-rt} X_t^{(1)} dt + (1-\zeta) \int_0^{\infty} e^{-rt} X_t^{(0)} dt \\
&= \zeta B + (1-\zeta)B = B.
\end{aligned}$$

Furthermore, for $\zeta \in (0,1)$, the discounted flow utility that $X^{(\zeta)}$ offers player i at each time t is defined and equal to $e^{-\delta^i t}u(\zeta X_t^{(1)} + (1-\zeta)X_t^{(0)})$. By the strict concavity of $u(\cdot)$, this is strictly greater than the ζ-mixture of the discounted flow utilities offered by $X^{(0)}$ and $X^{(1)}$, i.e. $e^{-\delta_i t}(\zeta u(X_t^{(1)}) + (1-\zeta)u(X_t^{(0)}))$.

Thus X^{ζ} offers a payoff profile that is weakly Pareto-superior to $\zeta U^1 + (1-\zeta)U^0$. It follows that the frontier of efficient payoffs cannot exhibit any (even weak) convexities.

Because the frontier of efficient payoffs is concave, an efficient collective schedule X must maximize

$$U^a(X) \equiv aU^H(X) + (1-a)U^L(X)$$

for some payoff weight $a \in [0,1]$. In particular, given efficient collective schedule X, the corresponding U^a cannot be increased by shifting resources between time 0 and any other time t. Given that $X_{(\cdot)}$ is right-continuous, this implies

$$X_0^{-\gamma} = e^{rt}\big(ae^{-\delta^H t} + (1-a)e^{-\delta^L t}\big)X_t^{-\gamma} \quad \forall t. \tag{237}$$

So X is optimal according to time preference factor

$$\beta_t^{(a)} = ae^{-\delta^H t} + (1-a)e^{-\delta^L t}, \tag{238}$$

or time preference rate

$$\delta_t^{(a)} = -\frac{\dot{\beta}_t^{(a)}}{\beta_t^{(a)}} = \frac{a\delta^H e^{-\delta^H t} + (1-a)\delta^L e^{-\delta^L t}}{ae^{-\delta^H t} + (1-a)e^{-\delta^L t}}. \tag{239}$$

As we can see, $\delta_0^{(a)} = a\delta^H + (1-a)\delta^L$. Therefore a is not only the weight placed

on H's payoff, but also the weight placed on her time preference in determining the starting time preference rate.

Given payoff weight a, let $w_t^{(a)}$ denote the weight placed on H's time preference rate at t, such that

$$\delta_t^{(a)} = w_t^{(a)}\delta^H + (1 - w_t^{(a)})\delta^L. \tag{240}$$

Substituting (239) into (240) and rearranging, we have

$$w_t^{(a)} = \frac{a}{a + (1-a)e^{(\delta^H - \delta^L)t}}.$$

Observe that

$$w_0^{(a)} = a. \tag{241}$$

Fixing payoff weight a, the resulting schedule is not time-consistent, because the resulting time preference rate (239) is not constant. Upon reaching time $s > 0$, $aU^H + (1-a)U^L$ is maximized across times $t \geq s$ by allocating the collective budget according to time preference rate schedule δ_{t-s}, as prescribed, not according to a fixed δ_s.

However, if upon reaching s we instead place weight

$$\tilde{a} \equiv w_s^{(a)}$$

on H's future payoff (and $1 - \tilde{a}$ on L's), the resulting time preference rate schedule

is the same across $t \geq s$ as that prescribed at time 0 using payoff weight a. That is,

$$\delta_{t-s}^{(\tilde{a})} = \delta_t^{(a)} \quad \forall s \geq t.$$

We can see this by substituting $w_s^{(a)}$ for a into (238), simplifying, and differentiating:

$$\begin{aligned}
\beta_{t-s}^{(\tilde{a})} &= \frac{a \cdot e^{-\delta^H(t-s)}}{a + (1-a)e^{(\delta^H - \delta^L)s}} + \frac{(1-a)e^{(\delta^H-\delta^L)s} \cdot e^{-\delta^L(t-s)}}{a + (1-a)e^{(\delta^H-\delta^L)s}} \\
&= \frac{e^{\delta^H s}}{a + (1-a)e^{(\delta^H-\delta^L)s}}\left(ae^{-\delta^H t} + (1-a)e^{-\delta^L t}\right) \\
\implies \delta_{t-s}^{(\tilde{a})} &= -\frac{\dot{\beta}_{t-s}^{(\tilde{a})}}{\beta_{t-s}^{(\tilde{a})}} = \frac{a\delta^H e^{-\delta^H t} + (1-a)\delta^L e^{-\delta^L t}}{ae^{-\delta^H t} + (1-a)e^{-\delta^L t}} = \delta_t^{(a)}.
\end{aligned}$$

Let $X^{(a)}$ now denote the efficient collective schedule implied by payoff weight a. It is unique by the strict concavity of the feasible payoff set. It is also $\mathcal{C}^1$ in a, as can be seeing by rearranging and integrating (237) to get

$$\int_0^\infty e^{\frac{r-r\gamma}{\gamma}} X_0^{(a)} \frac{1}{ae^{-\delta^H t} + (1-a)e^{-\delta^L t}} dt - B = 0 \tag{242}$$

and applying the implicit function theorem to find that $X_0^{(a)}$, and thus $X_t^{(a)}$ for each t, is $\mathcal{C}^1$ in a.

Let $\chi^{(a)}$ denote the corresponding collective schedule defined by

$$\chi_t^{(a)} \equiv X_t^{(a)}/B.$$

Likewise, for $b^H \in [0,1]$, let $\chi^*[b^H]$ denote the collective schedule defined by

$$\chi_t^*[b^H] \equiv X_t^*[Bb^H, B(1-b^H)]/B.$$

Given b^H, by the homotheticity of $U(\cdot)$, $X^{(a)}$ is [weakly] Pareto superior to X^* iff $\chi^{(a)}$ is [weakly] Pareto superior to $\chi^*[b^H]$.

Given any $\chi^{(a)}$ for $a \in (0,1)$, there is an interval $(\underline{b}^H, \overline{b}^H) \subset (0,1)$ such that $\chi^{(a)}$ is a strict Pareto improvement on $\chi^*[b^H]$ iff $b^H \in (\underline{b}^H, \overline{b}^H)$. This follows from the inefficiency of $\chi^*[b^H]$ for $b^H \in (0,1)$ and the facts that

i. $U\big(\chi^*[0]\big) = U\big(\chi^{(0)}\big)$,

ii. $U\big(\chi^*[1]\big) = U\big(\chi^{(1)}\big)$,

iii. $U^H\big(\chi^*[b^H]\big)$ is continuous and strictly increasing (in fact $\mathcal{C}^1$ with a positive derivative) in b^H, and

iv. $U^L\big(\chi^*[b^H]\big)$ is continuous and strictly decreasing (in fact $\mathcal{C}^1$ with a negative derivative) in b^H.

We can thus define

$$\underline{b}^H(a) \equiv \underset{b^H}{\text{argmin}} : U^L\big(\chi^{(a)}\big) \geq U^L\big(\chi^*[b^H]\big),$$
$$\overline{b}^H(a) \equiv \underset{b^H}{\text{argmax}} : U^H\big(\chi^{(a)}\big) \geq U^H\big(\chi^*[b^H]\big),$$

with $\overline{b}^H(a) > \underline{b}^H(a)$. Recall that $X^{(a)}$, and thus $U^i(\chi^{(a)})$, are $\mathcal{C}^1$ in a. By the implicit function theorem and (iii)-(iv), $\underline{b}^H(a)$ and $\overline{b}^H(a)$ are also $\mathcal{C}^1$ in a, with negative derivatives.

By construction, $(\underline{b}^H(a), \overline{b}^H(a))$ is the range of initial impatient budget shares b^H such that $\chi^{(a)}$ is strictly Pareto superior to $\chi^*[b^H]$. As noted above, by homotheticity of $U(\cdot)$, it is also the b^H-range such that $X^{(a)}$ is Pareto superior to X^*. So both parties

weakly prefer $X^{(a)}$ to X^* at $t = 0$ iff

$$b^H \in \big(\underline{b}^H(a), \overline{b}^H(a)\big). \tag{243}$$

More generally, given a schedule x, it holds at all t that both parties weakly prefer the forward-looking collective schedule $X^{(a)}_{[t,\infty)}$ to the forward-looking Stackelberg collective schedule $X^*[t, \infty, B^H(x_{|t}), B^L(x_{|t})]$ iff

$$b^H(x_{|t}) \in \big(\underline{b}^H\big(w_t^{(a)}\big), \overline{b}^H\big(w_t^{(a)}\big)\big) \ \ \forall t. \tag{244}$$

Observe that $\frac{\partial}{\partial t} w_t^{(a)} < 0$ and thus that $\frac{\partial}{\partial t} \underline{b}^H\big(w_t^{(a)}\big) < 0$ and $\frac{\partial}{\partial t} \overline{b}^H\big(w_t^{(a)}\big) < 0$.

We will now show that, for any a satisfying (243), there is a schedule x such that $X = X^{(a)}$ and (244) holds.

A schedule everywhere superior to forward-looking Stackelberg

Observe that $\underline{b}^H\big(w_t^{(a)}\big)$ and $\overline{b}^H\big(w_t^{(a)}\big)$ are $\mathcal{C}^1$ in t for any payoff weight a. We can therefore define a differentiable path of budget shares for H, $\{\tilde{b}_t^H\}$, by

$$\begin{aligned} \tilde{b}_0^H &= b^H; \\ \frac{d}{dt}\tilde{b}_t^H &= \max\Big(0, \frac{1+p-2p_t}{1-p}\Big) \max\Big(f_1(\tilde{b}_t^H, t), f_2(\tilde{b}_t^H, t)\Big) + \min\Big(2\frac{p_t - p}{1-p}, 1\Big) f_2(\tilde{b}_t^H, t), \end{aligned} \tag{245}$$

where

$$f_1(\tilde{b}_t^H, t) \equiv \frac{(\tilde{b}_t^H - \underline{b}^H)\left(\frac{d}{dt}\overline{b}^H(w_t^{(a)}) - \frac{d}{dt}\underline{b}^H(w_t^{(a)})\right)}{\overline{b}^H(w_t^{(a)}) - \underline{b}^H(w_t^{(a)})}, \quad f_2(\tilde{b}_t^H, t) \equiv (\tilde{b}_t^H - 1)\frac{X_t^{(a)}}{B(X_{|t}^{(a)})},$$

$$p \equiv \frac{b^H - \underline{b}^H(a)}{\overline{b}^H(a) - \underline{b}^H(a)} \in (0,1), \qquad p_t \equiv \frac{\tilde{b}_t^H - \underline{b}^H\left(w_t^{(a)}\right)}{\overline{b}^H\left(w_t^{(a)}\right) - \underline{b}^H\left(w_t^{(a)}\right)}.$$

Because the right-hand side of (245) is continuous in t (fixing $\tilde{b}_t^H$) and Lipschitz continuous in $\tilde{b}_t^H$ across $[0,T]$ for all T, $\tilde{b}_T^H$ is defined and $\mathcal{C}^1$ for all T by the Picard–Lindelöf theorem.

Recall that $p_0 = p$ by (241), and observe that $p_{(\cdot)}$ is differentiable. We will show that $p_t \in (0,1)$ for all t.

First, observe that $\frac{d}{dt}\tilde{b}_t^H$ linearly interpolates, in p_t, from $\max(f_1, f_2)$ at $p_t \leq p$ to f_2 at $p_t \geq \frac{1+p}{2}$. Also, $\frac{d}{dt}p_t$ strictly increases in $\frac{d}{dt}\tilde{b}_t^H$, and when $\frac{d}{dt}\tilde{b}_t^H = f_1(\tilde{b}_t^H, t)$, $\frac{d}{dt}p_t = 0$. So if $p_t \leq p$, $\frac{d}{dt}p_t \geq 0$. This establishes that we never reach a t with $p_t < p$.

Next, suppose by contradiction that $p_{t^*} = 1$ for some t^*. Let t^* denote the minimum such time, noting that a minimum exists because $p_{(\cdot)}$ is continuous and $[0,\infty)$ is closed below. Then $\tilde{b}_{t^*}^H / \overline{b}^H\left(w_{t^*}^{(a)}\right) = 1$. Since p_t is continuous and $p_0 < 1$, there is an $s \in (0, t^*)$ such that $p_t \in [\frac{1+p}{2}, 1)\ \forall t \in [s, t^*]$, and therefore also such that $\tilde{b}_t^H / \overline{b}^H\left(w_t^{(a)}\right) < 1$ throughout this interval. However,

$$\frac{d}{dt}\frac{\tilde{b}_t^H}{\overline{b}^H(w_t^{(a)})} \leq 0 \ \ \forall t \in [s, t^*].$$

This is because the proportional, and thus the absolute, growth rate of $\tilde{b}^H(x_{|t}) / \overline{b}^H(w_t^{(a)})$ is non-positive iff

$$\frac{\frac{d}{dt}\overline{b}^H\left(w_t^{(a)}\right)}{\overline{b}^H\left(w_t^{(a)}\right)} \geq \frac{\frac{d}{dt}b^H(x_{|t})}{b^H(x_{|t})}.$$

By (245) and the fact that $\tilde{b}^H(x_{|t}) \leq \overline{b}^H(w_t^{(a)})$ for $t \in [s, t^*]$,

$$\frac{\frac{d}{dt}\tilde{b}^H(x_{|t})}{\tilde{b}^H(x_{|t})} = \frac{\big(\tilde{b}^H(x_{|t}) - 1\big)X_t^{(a)}/B(X_{|t}^{(a)})}{b^H(x_{|t})} \leq \frac{\big(\overline{b}^H\big(w_t^{(a)}\big) - 1\big)X_t^{(a)}/B(X_{|t}^{(a)})}{\overline{b}^H\big(w_t^{(a)}\big)} \;\; \forall t \in [s, t^*].$$

The proof is thus completed by showing that

$$\frac{\frac{d}{dt}\overline{b}^H\big(w_t^{(a)}\big)}{\overline{b}^H\big(w_t^{(a)}\big)} \geq \frac{\big(\overline{b}^H(w_t^{(a)}) - 1\big)X_t^{(a)}/B(X_{|t}^{(a)})}{\overline{b}^H(w_t^{(a)})} \;\; \forall t \in [s, t^*]. \tag{246}$$

Suppose by contradiction that (246) fails for some $t \in [s, t^*]$. The right-hand side equals $(\frac{d}{dt}b_t^H)/b_t^H$ given $b_t^H = \overline{b}^H(w_t^{(a)})$ and $x_t^H = X_t^{(a)}$. But $b_t^H = \overline{b}^H(w_t^{(a)})$ implies that H is indifferent between the forward-looking efficient and Stackelberg schedules:

$$U^H\big(X_{[t,\infty)}^{(a)}\big) = U^H\big(X^*[t, \infty, B^H(x_{|t}), B^L(x_{|t})]\big). \tag{247}$$

$b_t^H > 0$ and the failure of (246) then imply that there is an $\epsilon > 0$ such that it is feasible to set $x_s^H = X_s^{(a)}$ for $s \in [t, t+\epsilon)$, and such that if this obtains, $\overline{b}^H$ falls by a larger proportion than b^H across the interval. We then have $\underline{b}^H\big(w_{t+\epsilon}^{(a)}\big) < b_{t+\epsilon}^H$ and thus

$$\begin{aligned} &U^H\big(X_{[t+\epsilon,\infty)}^{(a)}\big) < U^H\big(X^*[t+\epsilon, \infty, B^H(x_{|t+\epsilon}), B^L(x_{|t+\epsilon})]\big) \implies \\ &U^H\big(X_{[t,\infty)}^{(a)}\big) < U^H\big(X_{[t,t+\epsilon)}^{(a)}\big) + e^{-\delta^H\epsilon}U^H\big(X^*[t+\epsilon, \infty, B^H(x_{|t+\epsilon}), B^L(x_{|t+\epsilon})]\big). \end{aligned} \tag{248}$$

But this is impossible, since the schedule across the right-hand side of (248)—$x_s^H = X_s^{(a)}, x_s^L = 0$ for $s \in [t, t+\epsilon)$ followed by a Stackelberg schedule—is an IC-polarized schedule from t onward, and H's favorite such schedule is the Stackelberg schedule itself from t onward (see Appendix A.5, "Stackelberg schedules", the $\bar{t} = \infty$

case), whose payoff for H equals the left-hand side of (248) by (247). So (246) holds.

So, given the path of impatient budget shares $\{\tilde{b}_t^H\}$ characterized by (245), $p_t \in (0,1)\ \forall t$.

We will now show that there is a unique schedule $x^{(a)}$ such that $x^{H(a)} + x^{L(a)} = X^{(a)}$ and $b^H(x^{(a)}_{|t}) = \tilde{b}_t^H\ \forall t$, and confirm that $x^{(a)}$ is (i) feasible in the sense that $x_t^{i(a)} \geq 0\ \forall i, t$ and (ii) continuous. We will not need to further verify that $x^{(a)}$ does not violate either player's budget constraint, since we have already established that $\tilde{b}_t^H \in (0,1)\ \forall t$.

Given a schedule $x^{(a)}$ with $x^{H(a)} + x^{L(a)} = X^{(a)}$, $b^H(x^{(a)}_{|t}) = \tilde{b}_t^H\ \forall t$ iff

$$\frac{d}{dt}\tilde{b}_t^H = \frac{d}{dt}b^H(x^{(a)}_{|t}) = \frac{\tilde{b}_t^H X_t^{(a)} - x_t^{H(a)}}{B(X^{(a)}_{|t})}$$

$$\implies \frac{x_t^{H(a)}}{X_t^{(a)}} = \tilde{b}_t^H - \frac{d}{dt}\tilde{b}_t^H \cdot \frac{B(X^{(a)}_{|t})}{X_t^{(a)}} \tag{249}$$

(with $x_t^{L(a)}/X_t^{(a)} = 1 - x_t^{H(a)}/X_t^{(a)}$).

$x^{(a)}$ is thus well defined. Furthermore (249) > 1 at t iff $\frac{d}{dt}\tilde{b}_t^H < f_2(\tilde{b}_t^H, t)$, and (249) < 0 at t only if $\frac{d}{dt}\tilde{b}_t^H < 0$. Both are impossible, by (245) and the non-positivity of f_1 and f_2. So $x^{(a)}$ is feasible. Finally, because $\tilde{b}_t^H$ is $\mathcal{C}^1$ in t, $x^{(a)}$ is continuous.

This completes the construction of a feasible schedule $x^{(a)}$ that is continuous and satisfies (244).

Constructing a grid and equilibrium

Given a grid G and a payoff weight a satisfying (243), let σ^* be a polarized equilibrium, and let $\tilde{\sigma}$ be the strategy profile of the game played on G with

$$\tilde{\sigma}^i(h_{|t}) = \begin{cases} \tau'(t), & h_{|t} = h^-_{|t}(\tilde{\sigma}); \\ x^{i(a)}_{[t,\tau'(t))}, & h_{|t} = h^+_{|t}(\tilde{\sigma}), \\ \sigma^{i*}(h_{|t}) & \text{otherwise.} \end{cases} \tag{250}$$

Note that any node reached on the path of $\tilde{\sigma}$ must occur at a grid point $t \in G$, and so be of type 1 or type 2.

We will now construct a grid G such that σ^* is an equilibrium of the game played on G.

Let $\tau_0 = 0$, and define τ_k for $k > 1$ recursively as follows.

Let $\hat{X}[t, \epsilon]$ be the truncated schedule across $[t, \infty)$ with

$$\begin{aligned} \hat{X}_s[t,\epsilon] \equiv X^*_{t+\epsilon}[t+\epsilon, \infty, \hat{B}^H[t,\epsilon], \hat{B}^H[t,\epsilon]]e^{(r-\alpha^H)(s-(t+\epsilon))}, \quad & s \in [t, t+\epsilon); \\ X^*[t+\epsilon, \infty, \hat{B}^H[t,\epsilon], \hat{B}^L[t,\epsilon]], \quad & s \geq t+\epsilon, \end{aligned}$$

where

$$\hat{B}^i[t,\epsilon] \equiv \Big(B^i\big(x^{(a)}_{|t}\big) + \int_t^{t+\epsilon} e^{-r(s-t)} x_s^{-i(a)} ds\Big) e^{r\epsilon} \quad \forall i.$$

To interpret $\hat{X}[t, \epsilon]$, suppose that $x_{|t} = x^{(a)}_{|t}$. Suppose then that, at t, each i receives a budget-increase equal to the present value of the resources allocated by $-i$ across $[t, t+\epsilon)$ according to $x^{-i(a)}$; invests the entirety of the newly increased B^i_t until $t+\epsilon$

(so that their budget at $t+\epsilon$ is then $\hat{B}^i[t,\epsilon]$); and subsequently implements the Stackelberg schedule. Then

$$X_{t+\epsilon} = X^*_{t+\epsilon}[t+\epsilon, \infty, \hat{B}^H[t,\epsilon], \hat{B}^L[t,\epsilon]].$$

$\hat{X}_{[t,t+\epsilon)}[t,\epsilon]$ is then the truncated collective schedule obtained across $[t, t+\epsilon)$ by positing that collective spending grows at rate $r - \alpha^H$ across $[t, t+\epsilon)$ (as it does just after $t+\epsilon$) and is continuous at $t+\epsilon$. $\hat{X}[t,\epsilon]$ is of course infeasible given initial budgets $\{B^i(x^{(a)}_{|t})\}$, for any $\epsilon > 0$. In particular, observe that for all $s \geq t$, $\hat{X}_s[t,0] = X^*_s[t,\infty, B^H(x^{(a)}_t), B^L(x^{(a)}_t)]$, and $\hat{X}_s[t,\epsilon]$ is $\mathcal{C}^1$ in ϵ with a derivative strictly positive for all $\epsilon \geq 0$. Observe also that $\hat{X}$ is Stackelberg with respect to the present-value budgets it allocates from player.

Then let

$$\epsilon^i(t) \equiv \epsilon > 0 : U^i\big(X^{(a)}_{[t,\infty)}\big) = U^i\big(\hat{X}[t,\epsilon]\big),$$
$$\epsilon(t) \equiv \min(\epsilon^H(t), \epsilon^L(t)).$$

$\epsilon^i(t)$ exists and is unique for each i because

$$U^i\big(\hat{X}[t,0]\big) = U^i\big(X^*[t,\infty, B^H(x^{(a)}_{|t}), B^L(x^{(a)}_{|t})]\big) < U^i\big(X^{(a)}_{[t,\infty)}\big);$$

$\lim_{\epsilon\to\infty} U^i(\hat{X}[t,\epsilon])$ is the supremum feasible payoff (∞ if $\gamma \leq 1$, 0 if $\gamma > 1$); and $U^i(\hat{X}[t,\epsilon])$ is continuous and monotonic in ϵ.

Let $\tau_k = \tau_{k-1} + \epsilon(\tau_{k-1})$.

We must establish that $\lim_{k\to\infty} \tau_k = \infty$, so that the grid is locally finite and the game is well-defined. This follows from the facts that, if not, then $\lim_{k\to\infty} \tau_k = \overline{\tau}$

for some $\overline{\tau} < \infty$, by the monotone convergence theorem, and thus $\epsilon^i(\tau_k) \to 0$ for some i; that $\epsilon^i(t)$ is continuous in t for each i, by the implicit function theorem; and $\epsilon^i(\overline{\tau}) > 0$ for each i.

To establish that $\tilde{\sigma}$ is an equilibrium of the game played on grid $G = \{\tau_k\}_{k \in \mathbb{N}}$, we will show that no player profits by deviating at any node. Given a node τ, we will let $\hat{X} \equiv \hat{X}[\tau, \tau' - \tau]$.

If $h_{|t}$ is off the path of $\tilde{\sigma}$, this follows from the fact that σ^* is an SPE.

If $h_{|\tau}$ is on path and of type 1, a deviation by H induces an IC-polarized subsequent schedule, which she must weakly disprefer to $X^*[t, \infty, B^H_\tau, B^L_\tau]$ (see Appendix A.5); and since $X^*_s[t, \infty, B^H_\tau, B^L_\tau] < \hat{X}_s \ \forall s \geq \tau$, she prefers the latter schedule. Since a deviation by L to an announcement $\xi^L < \tau'$ will be followed by polarized equilibrium behavior from τ onward, it follows from Appendix A.5: Inductive step: Polarization of equilibrium schedules given placement of first announcement (the $\hat{\xi} = \xi^H < \xi^L$ case) that L weakly prefers the schedule implemented by σ^* given $\xi^L = \tau'$ to that implemented by σ^* given any other ξ^L. This schedule likewise exhibits lower spending than $\hat{X}$ across $[\tau, \infty)$ in all three cases (open-loop, quasi-Stackelberg, and Stackelberg).

If $h_{|\tau}$ is on path and of type 2, first observe that

$$X^*_{\tau'}[\tau', \infty, B^H(x^{(a)}_{|\tau'}), B^L(x^{(a)}_{|\tau'})] > X^{(a)}_{\tau'}. \tag{251}$$

If (251) failed, then, because the growth rate of $X^{(a)}$ is everywhere within $(r - \alpha^H, r - \alpha^L)$ by (238), and because $X^{(a)}$ cannot lie everywhere above X^* after τ' given the same initial budgets since both collective schedules are budget-exhausting,

there would have to be a $\bar{t} > t^*[\tau', \infty, B^H(x^{(a)}_{|\tau'}), B^L(x^{(a)}_{|\tau'})]$ such that

$$X^*_t[\tau', \infty, B^H(x^{(a)}_{|\tau'}), B^L(x^{(a)}_{|\tau'})] < / = / > X^{(a)}_t, \quad t < / = / > \bar{t}.$$

$X^{(a)}$ could thus be constructed by shifting resources from times after $\bar{t}$ to times before $\bar{t}$. Since all of the latter offer weakly lower δ^L-discounted marginal utility than all of the former, $X^{(a)}_{[\tau', \infty)}$ is then not a Pareto improvement on $X^*[\tau', \infty, B^H(x^{(a)}_{|\tau'}), B^L(x^{(a)}_{|\tau'})]$, as we know it to be by construction.

Next, observe that if X is a collective schedule following τ that i can implement by deviating at $h_{|\tau}$, i can also implement X given

$$\begin{aligned} B^i_\tau = \tilde{B}^i_\tau &\equiv B^i(x^{(a)}_{|\tau}) + \int_\tau^{\tau'} e^{-r(t-\tau)} x_t^{-i(a)} dt, \\ B^{-i}_\tau = \tilde{B}^{-i}_\tau &\equiv B^{-i}(x^{(a)}_{|\tau}) - \int_\tau^{\tau'} e^{-r(t-\tau)} x_t^{-i(a)} dt, \\ x_t^{-i} &= 0, \quad t \in [\tau, \tau') \end{aligned}$$

(and polarized equilibrium behavior from τ' onward) by reserving the resources "transferred" from the other player for spending before τ' as prescribed by $x^{-i(a)}$.

Under these circumstances, with $i = H$, any schedule H implements is an IC-polarized allocation of $\{\tilde{B}^i\}$. It follows from (251) that

$$\tilde{B}^H < \int_\tau^\infty e^{-r(t-\tau)} \hat{X}^H_t dt,$$

and so the same also holds for L. To summarize, therefore, H prefers $\hat{X}$ to $X^*[\tau, \infty, \tilde{B}^H, \tilde{B}^L]$, as both are Stackelberg but the former uses larger budgets for both players (and $U^H(X^*[B^H, B^L])$ is increasing in both arguments); weakly prefers $X^*[\tau, \infty, \tilde{B}^H, \tilde{B}^L]$ to any IC-polarized allocation of $\{\tilde{B}^i\}$; and weakly prefers an IC-

polarized allocation of $\{\tilde{B}^i\}$ to any schedule she can implement by deviating at τ.

With $i = L$, he maximizes his payoff by setting a δ^L-optimal schedule across $[\tau, \tau')$ with some fraction of his $\tilde{B}^L_\tau$ and allocating the rest to the Stackelberg schedule that will obtain following τ'. (A maximal payoff exists by the extreme value theorem, since U^L is continuous in the fraction of $\tilde{B}^L_\tau$ allocated before τ'.) Whatever fraction he chooses, the resulting schedule following τ' is Stackelberg with an initial budget for H of $\tilde{B}^H e^{r(\tau'-\tau)} < \hat{B}^H$ and an initial budget for L weakly less than $\hat{B}^L$ (with equality only if L spends nothing before τ').

Since $U^L(X^*[B^H, B^L])$ increases in both arguments (see Appendix A.5: Final period: Unique equilibrium schedule given placement of first announcement), an optimal deviation by L produces a schedule X with

$$U^L(X_{[\tau',\infty)}) < U^L(\hat{X}_{[\tau',\infty)}).$$

The cited subsection also establishes that an optimal division of $\tilde{B}^L_\tau$ across τ' must yield $\lim_{t\to\tau'^+} X_t \leq X_{\tau'}$, and that in general $X^*_t[t, \infty, B^H_t, B^L_t]$ increases in B^i_t for both i. Since $\hat{B}^i$ must be weakly greater than $B^i(x_{|\tau'})$ for both i, $\lim_{t\to\tau'^+} X_t \leq \hat{X}_{\tau'}$. Finally, since to be δ^L-optimal X must grow at rate $r - \alpha^L$ across $[\tau, \tau')$, and since $\hat{X}$ grows at the slower rate $r - \alpha^H$ across this interval, $\hat{X}_s > X_s \; \forall s \in [\tau, \tau')$ and thus

$$U^L(X_{[\tau,\tau')}) < U^L(\hat{X}_{[\tau,\tau')}).$$

We have shown that σ^* is an equilibrium on any grid G with $\tau' \leq \tau + \epsilon(\tau) \; \forall \tau \in G$.

To find an admissible grid sequence such that σ^* is an equilibrium of every grid

in the sequence, let $f(t) \equiv \{f_k(t)\}_{k\in\mathbb{N}}$ be the sequence defined recursively by

$$f_0(t) = t, \;\; f_k(t) = f_{k-1}(t) + \epsilon(f_{k-1}(t)) \;\; (k > 0), \tag{252}$$

so that $G = f(0)$. Then let $G^{(0)} = G$, and let

$$G^{(n)} = G^{(n-1)} \cup \bigcup_{m=0}^{\infty} f\Big(\frac{m}{2^{n-1}}\Big), \;\; n > 0.$$

Intermediate equilibrium payoffs

So far we have established that any payoff profile $\overline{U}$ that is (i) efficient and (ii) strictly Pareto superior to $U(X^*)$ is an equilibrium payoff profile. We will now show that any feasible payoff profile U that is Pareto superior to $U(X^*)$ is an equilibrium payoff profile.

Let $[\underline{a}, \overline{a}]$ be the range of values of a such that $X^{(a)}$ is Pareto superior to X^*. Observe that, since $\chi_t^{(a)}$ is continuous in a for all t and nowhere constant in a (see (242)), $U^i\big(X^{(a)}\big)$ is continuous and strictly monotonic in a for each i.

Define $f^{(a)}(t)$ as in (252), but with $\epsilon(\cdot)$ defined with respect to a.

Suppose U is efficient but only weakly Pareto superior to $U(X^*)$, with $U^i = U^i(X^*)$. Let

$$a^i \in \{\underline{a}, \overline{a}\} \equiv a : U^i\big(X^{(a)}\big) = U^i\big(X^*\big),$$

and recall that for any $a \in (\underline{a}, \overline{a})$, there a grid G and a strategy profile σ such that σ is an equilibrium of the game given grid G and $X(\sigma) = X^{(a)}$.

We can therefore define a monotonic sequence $\{a_n\}$ with $a_0 \in (\underline{a}, \overline{a})$ and $a_n \to a^i$,

and a corresponding grid sequence with $G^{(0)} = f^{(a_0)}(0)$ and

$$G^{(n)} = \bigcup_{k=0}^{\infty} f^{(a_n)}\big(\tau_k^{(n-1)}\big) \cup \bigcup_{m=0}^{\infty} f^{(a_n)}\Big(\frac{m}{2^{n-1}}\Big), \quad n > 0.$$

The first union ensures that $G^{(n)} \supset G^{(n-1)}$. The second ensures that $\cup_{n=0}^{\infty} G^{(n)}$ is dense. The inclusion of all elements of $f^{(a_n)}(\tau)$ for $\tau \in G^{(n)}$ ensures that, for each n, there is a strategy profile $\sigma^{(n)}$ that is an equilibrium of the game played on $G^{(n)}$ and such that $X(\sigma^{(n)}) = X^{(a_n)}$. Since $X_t^{(a_n)} \to X_t^{(a^i)}$ for all t, $X^{(a^i)}$ is an equilibrium collective schedule. Since $U(X^{(a_n)}) \to U = U(X^{(a^i)})$, U is an equilibrium payoff.

Suppose U is Pareto superior to $U(X^*)$ but inefficient. Then there is an efficient payoff profile $\overline{U}$ strictly Pareto superior to U (and so also to $U(X^*)$) and a $C \in (0,1)$ such that

$$U = \overline{U}[C] \equiv \begin{cases} C^{1-\gamma}\overline{U}, & \gamma \neq 1; \\ \big(\overline{U}^H + \ln(C), \overline{U}^L + \ln(C)\big), & \gamma = 1. \end{cases}$$

Let $a \in (0,1)$ be such that $U(X^{(a)}) = \overline{U}$.

Let σ^* be a polarized equilibrium, and for $c \in (C,1)$, let $\sigma^{(c)}$ be the strategy profile with

$$\sigma^i(h_{|t}) = \begin{cases} \tau'(t), & h_{|t} = h_{|t}^-(\sigma); \\ \big\{\frac{B^i}{\tau_1}(1-c)e^{rs}\big\}_{s\in[0,\tau_1)}, & t = 0 \text{ and } h_{|t} = h_{|t}^+(\sigma); \\ ce^{r\tau_1} x^{i(a)}_{[t-\tau_1,\tau'(t)-\tau_1)}, & t > 0 \text{ and } h_{|t} = h_{|t}^+(\sigma); \\ \sigma^{i*}(h_{|t}), & \text{otherwise.} \end{cases}$$

Given $\sigma^{(c)}$, at time 0, each i chooses announcement $\xi^i = \tau_1$ and then allocates fraction

$1-c$ of B^i uniformly across $[0,\tau_1)$. The players then play $\tilde{\sigma}$, as defined by (250), with initial time τ_1 in place of 0 and initial budgets $\{ce^{r\tau_1}B^i\}$ in place of $\{B^i\}$. We will now find a τ_1 such that $\sigma^{(c)}$ is an equilibrium of the game played on grid $G \equiv 0 \cup \big(\tau_1 + f^{(a)}(0)\big)$, where $\tau_1 + f^{(a)}(0)$ denotes the set $f^{(a)}(0)$ with τ_1 added to each element. Note that $f^{(a)}(0)$ constructs a grid on which $\tilde{\sigma}$ is an equilibrium given any pair of budgets in the same ratio as $\{B^i\}$, as $\{ce^{r\tau_1}B^i\}$ is.

It follows from the fact that $\tilde{\sigma}$ and σ^* are SPEs that σ is an equilibrium at nodes after time 0. To verify that σ is an equilibrium at 0 for sufficiently small τ_1, i's payoff to defecting at either the type 1 or the type 2 node at $t=0$ can be upper-bounded by the payoff i receives if the entire budget is allocated δ^i-optimally across $[0,\tau_1)$ and also (of course infeasibly) invested to τ_1 and subsequently Stackelberg:

$$\int_0^{\tau_1} e^{-\delta^i t} u\Big(\frac{B\alpha^i}{1-e^{-\alpha^i\tau_1}} e^{(r-\alpha^i)t}\Big)dt + e^{-\delta^i\tau_1} U^i\big(X^*[\tau_1,\infty, B^H e^{r\tau_1}, B^L e^{r\tau_1}]\big)$$
$$\to 0 + U^i\big(X^*[0,\infty,B^H,B^L]\big) = U^i(X^*)$$

as $\tau_1 \to 0^+$. $U^i(X(\sigma^{(c)}))$ converges to $\overline{U}^i[c]$ as $\tau_1 \to 0^+$. Since $c > C$, $\overline{U}^i[c] > \overline{U}^i[C] = U^i \geq U^i(X^*)$.

Given $c \in (C,1)$, let $\tau_1[c]$ be a time such that $\sigma^{(c)}$ is an equilibrium on any grid $G = 0 \cup \big(\tau_1 + f^{(a)}(0)\big)$ with $\tau_1 \leq \tau_1[c]$. Let $\{c_n\}$ be a decreasing sequence with $c_0 \in (C,1)$ and $c_n \to C$. We can then define a corresponding grid sequence with $G^{(0)} = 0 \cup \big(\tau_1[c_0] + f^{(a)}(0)\big)$ and

$$G^{(n)} = G^{(n-1)} \cup \bigcup_{k=0}^{\infty} \big(\tau_1^{(n)} + f^{(a)}\big(\tau_k^{(n-1)} - \tau_1^{(n)}\big)\big) \cup \bigcup_{m=0}^{\infty} f^{(a)}\Big(\frac{m}{2^{n-1}}\Big), \quad n>0,$$

where $\tau_1^{(n)} = \min(1/2^{n-1}, \tau_1[c_n])$. The first term ensures that $G^{(n)} \supset G^{(n-1)}$. The second ensures that $\sigma^{(c_n)}$ is an equilibrium of the on-path subgame beginning at $\tau_1^{(n)}$.

The third again ensures that $\cup_{n=0}^{\infty} G^{(n)}$ is dense.

$\{\sigma^{(c_n)}\}$ is thus a sequence of equilibria across an admissible grid sequence. Since $X_t(\sigma^{(c_n)}) \to CX_t^{(a)}$ $\forall t \neq 0$ (since $\tau_1^{(n)} \to 0$), $CX^{(a)}$ is an equilibrium collective schedule. Since $U\big(X(\sigma^{(c_n)})\big) \to U\big(CX^{(a)}\big) = U$, U is an equilibrium payoff.

A.7 Proof of Proposition 7

Existence and uniqueness of constrained-polarized equilibrium schedule

Let σ^* be a strategy profile in which, at each post-announcement node $h_{|t}$, for $s \in [t, \xi(h_{|t}))$ we have

$$\sigma_s^{H*}(h_{|t}) = B^H(h_{|t})\overline{\alpha}e^{(r-\overline{\alpha})(s-t)}, \tag{253}$$

$$\sigma_s^{L*}(h_{|t}) = \begin{cases} 0, & s < \overline{t}^*(h_{|t}); \\ \big(B^H(h_{|t})e^{(r-\overline{\alpha})(\overline{t}^*(h_{|t})-t)} + B^L(h_{|t})e^{r(\overline{t}^*(h_{|t})-t)}\big) & \\ \quad \cdot \alpha^L e^{(r-\alpha^L)(s-\overline{t}^*(h_{|t}))} - B^H(h_{|t})\overline{\alpha}e^{(r-\overline{\alpha})(s-t)}, & s \geq \overline{t}^*(h_{|t}), \end{cases} \tag{254}$$

where

$$\overline{t}^*(h_{|t}) \equiv t + \max\Big(0, \ln\Big(\frac{B^H(h_{|t})}{B^L(h_{|t})}\,\frac{\overline{\alpha} - \alpha^L}{\alpha^L}\Big)\Big/\overline{\alpha}\Big). \tag{255}$$

Observe that

i. by (253), $x^H((\sigma^{H*}, \sigma^L))$ is independent of σ^L;

ii. by (253) and (254), $X_s(\sigma^*)$ is the δ^L-optimal allocation of $B_{\overline{t}^*(h_{|t})}$ across $s \geq \overline{t}^*(h_{|t})$; and

iii. by (254) and (255), $x^L_{\overline{t}^*(h_{|t})}(\sigma^*) = 0$ if $\overline{t}^*(h_{|t}) > t$.

By (ii), $x^L(\sigma^*)$ exhausts L's budget in present-value terms. By (i) and (iii), any alternative schedule for L either would sub-optimally allocate his resources weakly after $\overline{t}^*(h_{|t})$ or would shift some of L's resources from weakly after $\overline{t}^*(h_{|t})$ to times before $\overline{t}^*(h_{|t})$ which, because $\alpha^L \leq \overline{\alpha}$ and X is continuous at $\overline{t}^*(h_{|t})$, offer weakly lower δ^L-discounted marginal utility. It follows that σ^{L*} is a best response to σ^{H*}, and that for any constrained-polarized equilibrium σ, $x(\sigma) = x(\sigma^*)$.

As with the unconstrained game of Proposition 5, the constrained game is continuous at infinity iff $\gamma < 1$. In this case, to prove that σ^{H*} is a best response to σ^{L*}, we can use the one-shot deviation principle. Let $h_{|t}$ be a post-announcement node and let $\xi \equiv \xi(h_{|t})$.

Let $Y_s \equiv e^{-rs}B_t^H$ denote the present value, as of time 0, of H's budget at s, and let it be written as a function of the history and/or strategy profile as B_t^H can. $\dot{Y}_s \leq 0$, since H has no outside income, and $\dot{Y}_s \geq \overline{\alpha}Y_s$ due to the spending constraint. Since x^H is right-continuous, any deviation by H at $h_{|t}$ to a strategy $\tilde{\sigma}^H$ with $\tilde{\sigma}_s^H(h_{|t}) < B_s^H\overline{\alpha}$ for some $s \in [t, \xi)$ raises $\dot{Y}$ from its lower bound for an interval following s, which raises Y_ξ and thus B_ξ^H. So a deviation by H shifts resources from some times in $[t, \xi)$ (a) to increases in B_ξ^H and possibly also (b) to later times in the $[t, \xi)$ interval. We will now show that both shifts lower H's payoff.

Regarding (b): let

$$\overline{\mathcal{T}} \equiv \{s \in [t, \xi) : \tilde{\sigma}_s^H(h_{|t}) > X_s(h_{|t}, \sigma^*)\}, \tag{256}$$
$$\underline{\mathcal{T}} \equiv \{s \in [t, \xi) : \tilde{\sigma}_s^H(h_{|t}) < X_s(h_{|t}, \sigma^*)\}$$

denote the set of times in $[t, \xi)$ at which H spends more (less) under σ^H than under

σ^{H*}. Let

$$\overline{Y} \equiv \int_{\overline{\tau}} e^{-rs}\Big(\tilde{\sigma}_s^H(h_{|t}) - B^H(h_{|t})e^{(r-\overline{\alpha})(s-t)}\Big)ds.$$

Let

$$t^{-1}(y) \equiv s : Y_t - Y_s(h_{|t}, \sigma^*) = y$$
$$\tilde{t}^{-1}(y) \equiv s : Y_t - Y_s(h_{|t}, (\tilde{\sigma}^H, \sigma^{L*})) = y.$$

Observe first that $t^{-1}(y)$ is defined for all but a countable number of $y \in [Y(h_{|t}), Y_\xi(h_{|t}, \sigma^*))$. If a value of y in this interval is undefined, there exist $t_1, t_2 > t_1$ such that

$$Y_{t_1}(h_{|t}, \sigma^*) = Y_{t_2}(h_{|t}, \sigma^*) = y,$$

so, by the monotonicity of $Y_{(\cdot)}$, $Y_s = y$ for all $s \in [t_1, t_2]$. Thus if there were an uncountable number of values of y in the interval for which $t^{-1}(y)$ were undefined, the interval would contain an uncountable number of disjoint intervals, which it does not (as can be seen e.g. from the fact that each interval contains a distinct rational, and the rationals are countable). Likewise, $\tilde{t}^{-1}(y)$ is defined almost everywhere for $(\tilde{\sigma}^H, \sigma^{L*})$. Now observe that, for any $y \in [0, \overline{Y}]$ for which $t^{-1}(y)$ and $\tilde{t}^{-1}(y)$ are defined,

$$\tilde{t}^{-1}(y) \geq t^{-1}(y).$$

Finally, the payoff gains for H achieved by any above-baseline expenditures before ξ implemented by $\tilde{\sigma}^H$ are bounded above by

$$\int_0^{\overline{Y}} e^{(r-\delta^H)\tilde{t}^{-1}(y)} u'\Big(X_{\tilde{t}^{-1}(y)}(h_{|t}, \sigma^*)\Big)dy, \tag{257}$$

and the payoff losses to the first $\overline{Y}$ units of below-baseline expenditures are bounded below by

$$\int_0^{\overline{Y}} e^{(r-\delta^H)t^{-1}(y)} u'\Big(X_{\tilde{t}^{-1}(y)}(h_{|t}, \sigma^*)\Big) dy. \tag{258}$$

(The integrals are defined because $t^{-1}(\cdot)$ and $\tilde{t}^{-1}(\cdot)$ are defined almost everywhere.) Since $\tilde{t}^{-1}(y) \geq t^{-1}(y)$ almost everywhere, and since $X_{(\cdot)}(h_{|t}, \sigma^*)$ always grows at a proportional rate weakly greater than $r - \alpha^H$, the expression in integral (257) is less than or equal to that in (258) for every value of y. So any reallocations within $[t, \xi)$ implemented by $\tilde{\sigma}^H$ from the σ^{H*} baseline cannot raise H's payoff.

Regarding (a): we will show that increases to B_ξ^H increase X_s for all $s \geq \xi$. Let $\tilde{h} \equiv h_\xi^+(h_{|t}, (\tilde{\sigma}^H, \sigma^{L*}))$ and $h^* \equiv h_\xi^+(h_{|t}, \sigma^*)$. Observe that $\overline{t}^*(h)$ is continuous and weakly increasing in $B^H(h)$, and that $x_s^H(h, \sigma^*)$ is proportional to $B^H(h)$ for all $s \geq \xi$.

If $\overline{t}^*(\tilde{h}) = t$, then $\overline{t}^*(h^*) = t$; the schedule implemented by σ^* from ξ onward is in either case the δ^L-optimal allocation of the collective budget B_ξ; and the increase to B_ξ^H proportionally increases X_s for all $s \geq \xi$.

If $\overline{t}^*(\tilde{h}) > t$, then the increase to B_ξ^H increases $\overline{t}^*$—i.e. $\overline{t}^*(\tilde{h}) > \overline{t}^*(h^*)$—unless $\overline{\alpha} = \alpha^L$ (in which case $\overline{t}^*(\tilde{h}) = \overline{t}^*(h^*) = \xi$ regardless of B_ξ^H). And $X_{\overline{t}^*(\tilde{h})}(\tilde{h}, \sigma^*) > X_{\overline{t}^*(\tilde{h})}(h^*, \sigma^*)$: if it were not, then the collective resources allocated across $[\overline{t}^*(\tilde{h}), \infty)$ would be weakly less following $\tilde{h}$ than following h^*, even while the resources allocated by H would be strictly greater, so L's spending (now beginning at $\overline{t}^*(\tilde{h}) \geq \overline{t}^*(h^*)$) would not exhaust his budget. Since $r - \overline{\alpha} \leq r - \alpha^L$, $X_s(\tilde{h}, \sigma^*) > X_s(h^*, \sigma^*)$ for all $s \in [\xi, \overline{t}^*(\tilde{h}))$. So again, the increase to B_ξ^H increases X_s for all $s \geq \xi$.

This proves that one-shot deviations do not raise H's payoff. If $\gamma < 1$, this completes the proof that σ^{H*} is a best response to σ^{L*}.

Suppose $\gamma \geq 1$. Given that post-announcement node $h_{|t}$ has been reached, denote H's continuation payoff to playing strategy σ^H from grid point $\tau > t$ onward after playing $\tilde{\sigma}^L$ at all nodes $h_{|s}$ with $s \in [t, \tau)$ by

$$C(\sigma^H, \tau) \equiv \int_{\tau}^{\infty} e^{-\delta^H (s-\tau))} u\big(X_s\big(h^-_{|\tau}(h_{|t}, \tilde{\sigma}), (\sigma^H, \sigma^{L*})\big)\big) ds,$$

and denote H's payoff to playing σ^H from $h_{|t}$ onward by $C(\sigma^H, t)$. By backward induction on one-shot deviations, deviation from σ^{H*} to $\tilde{\sigma}^H$ for only a finite length of time lowers H's payoff following $h_{|t}$. We therefore have

$$\begin{aligned} & C(\sigma^{H*}, t) > C(\tilde{\sigma}^H, t) + e^{-\delta^H(\tau - t)}\big(C(\sigma^{H*}, \tau) - C(\tilde{\sigma}^H, \tau)\big) \\ \implies & C(\tilde{\sigma}^H, \tau) - C(\sigma^{H*}, \tau) > e^{\delta^H(\tau - t)}\big(C(\tilde{\sigma}^H, t) - C(\sigma^{H*}, t)\big). \end{aligned}$$

If $\tilde{\sigma}^H$ is a profitable deviation for H at $h_{|t}$, the right-hand side of (235) is positive, so the difference in continuation payoffs as a function of $\tau - t$ must be "fast-growing", as defined to mean asymptotically bounded below by $c_0 e^{\delta^H(\tau - t)}$ for some constant $c_0 > 0$. $C(\tilde{\sigma}^H, \tau)$ can never exceed the continuation payoff for H at τ obtained if both parties invest all funds from t to τ and subsequently disburse them δ^H-optimally. (This is in fact a very loose upper bound, since it ignores the feasibility constraint imposed by the spending maximum.) This continuation payoff plateaus if $\gamma > 1$, and grows linearly in $\tau - t$ at absolute rate r if $\gamma = 1$. (See the payoff expression from Proposition 1, substituting $B(h_{|t})e^{r(\tau - t)}$ for B.) For the difference in continuation payoffs to be fast-growing, therefore, $C(\sigma^{H*}, \tau)$ must eventually be negative and its absolute value fast-growing.

But this is impossible, since $C(\sigma^{H*}, \tau)$ is bounded below by H's continuation payoff at τ in the event that she minimizes B^H_τ by spending as quickly as possible

until τ—so that

$$B_\tau^H = B_t^H e^{(r-\overline{\alpha})(\tau-t)}$$

—and then enjoys only $X_s = x_s^H(h_{|\tau}, \sigma^*)$ for $s \geq \tau$, so that we ignore contributions from L. By the homotheticity of $U(\cdot)$, this lower bound is proportional to $u(B_\tau^H)$. Thus if $\gamma = 1$, this lower bound is linear in τ, and if $\gamma > 1$, this lower bound is proportional to $-e^{(r-\overline{\alpha})(1-\gamma)\tau}$. By $\overline{\alpha} < \alpha^H$ and (111), $r < \overline{\alpha} \implies (r-\overline{\alpha})(1-\gamma) < \delta^H$.

So σ^{H*} is a best response to σ^{L*} given any value of γ.

Open-loop schedule

By (i), the proof that σ^{L*} is a best response to σ^{H*} also serves as a proof that, in the open-loop setting, $\overline{x}^L$ is a best response to $\overline{x}^H$. The proof that, given σ^{L*}, H weakly disprefers reallocations from σ^{H*} (at a node $h_{|t}$) implemented within the $[t, \xi)$ interval also, with $t = 0$ and $\xi = \infty$, serves as a proof that, in the open-loop setting, $\overline{x}^H$ is a best response to $\overline{x}^L$.

A.8 Proof of Proposition 8

Existence and uniqueness of constrained-polarized equilibrium schedule

Let σ^* be a strategy profile in which, at each post-announcement node $h_{|t}$, for $s \in [t, \xi(h_{|t}))$ we have

$$\sigma_s^{L*}(h_{|t}) = B^L(h_{|t})\underline{\alpha}e^{(r-\underline{\alpha})(s-t)}, \tag{259}$$

$$\sigma_s^{H*}(h_{|t}) = \begin{cases} \left(B^L(h_{|t})e^{(r-\underline{\alpha})(\underline{t}^*(h_{|t})-t)}\right)\underline{\alpha}e^{(r-\alpha^H)(s-\underline{t}^*(h_{|t}))} & \\ \quad -B^L(h_{|t})\underline{\alpha}e^{(r-\underline{\alpha})(s-t)}, & s < \underline{t}^*(h_{|t}), \\ 0, & s \geq \underline{t}^*(h_{|t}); \end{cases} \tag{260}$$

where $\underline{t}^*(h_{|t})$ uniquely satisfies

$$\alpha^H/b(h_{|t}) = \underline{\alpha}e^{(\alpha^H-\underline{\alpha})(\underline{t}^*-t)} + (\alpha^H - \underline{\alpha})e^{-\underline{\alpha}(\underline{t}^*-t)} \tag{261}$$

(or equals ∞ if $B^L(h_{|t}) = b(h_{|t}) = 0$, or if $\underline{\alpha} = \alpha^H$).

To verify that there is a $\underline{t}^*(h_{|t}) \geq t$ satisfying (261) if $b(h_{|t}) > 0$ and $\underline{\alpha} < \alpha^H$, observe that the right-hand side of (261) equals α^H at $\underline{t}^* = t$ and that, given $\alpha^H > \underline{\alpha}$, it rises without bound as $\underline{t}^* \to \infty$. To verify that $\underline{t}^*(h_{|t})$ is unique, observe that the derivative of the right-hand side of (261) equals 0 at $\underline{t}^* = t$ and is positive at $\underline{t}^* > t$.

Equality (261), and $\underline{t}^*(h_{|t}) = \infty$ given $b(h_{|t}) = 0$ or $\underline{\alpha} = \alpha^H$, is derived by setting

$$B^H(h_{|t}) = \int_t^{\underline{t}^*(h_{|t})} e^{-r(s-t)}\sigma_s^{H*}(h_{|t})ds,$$

so that H precisely exhausts her budget at $\underline{t}^*(h_{|t})$.

Observe that

i. by (259), $x^L((\sigma^H, \sigma^{L*}))$ is independent of σ^H;

ii. by (260), σ^* implements the δ^H-optimal allocation of resources spent across $[t, \underline{t}^*(h_{|t}))$; and

iii. by (260), $\lim_{s\to\underline{t}^{*-}(h_{|t})} x_t^H(\sigma^*) = 0$.

So any alternative schedule for H either would sub-optimally allocate her resources

weakly before $\underline{t}^*(h_{|t})$ or would shift some of her resources from weakly before $\underline{t}^*(h_{|t})$ to times on or after $\underline{t}^*(h_{|t})$ which, because $\alpha^H \geq \underline{\alpha}$ and X is continuous at $\underline{t}^*(h_{|t})$, offer weakly lower δ^H-discounted marginal utility. It follows that σ^{H*} is a best response to σ^{L*}, and that for any constrained-polarized equilibrium σ, $x(\sigma) = x(\sigma^*)$.

As with the unconstrained game of Proposition 5, the constrained game is continuous at infinity iff $\gamma < 1$. In this case, to prove that σ^{L*} is a best response to σ^{H*}, we can use the one-shot deviation principle. Let $h_{|t}$ be a post-announcement node and let $\xi \equiv \xi(h_{|t})$. By a proof precisely analogous to that in the case of a spending minimum (see Appendix A.8), deviation by L at $h_{|t}$ to a strategy implementing an alternative path of x^L until ξ must shift resources from all times on and after ξ to some times before ξ, and may also shift resources from later to earlier periods within $[t, \xi)$; and both effects lower L's payoff. If $\gamma < 1$, this completes the proof that σ^{L*} is a best response to σ^{H*}.

If $\gamma \geq 1$, the proof that σ^{L*} is a best response to σ^{H*} is, for the first two paragraphs, identical to the proof that σ^{H*} is a best response to σ^{L*} in the case of a spending minimum (where σ^* is defined as in Appendix A.8), with the L and H indices reversed. Here, to disprove that $C(\sigma^{L*}, \tau)$ is negative and fast-growing, let

$$B^i_\tau \equiv B^i(x_{|\tau}(h_{|t}, (\sigma^{H*}, \tilde{\sigma}^L))),$$

and observe that $C(\sigma^{L*}, \tau)$ is bounded below both by

- L's continuation payoff at τ if he allocates B^L_τ δ^L-optimally from τ onward in the absence of H, which is given by (134), and by
- L's continuation payoff at τ if H allocates B^H_τ δ^H-optimally from τ onward in the absence of L, which is given by $U_{\delta^L}(B^H_\tau, \alpha^H)$ as defined by (135).

As in the proof that a polarized equilibrium exists in the unconstrained setting (Appendix 5), this in turn implies for any $\gamma \geq 1$ that B_τ^L and B_τ^H (and thus also B_τ) are asymptotically bounded above by an expression that declines to zero quickly enough that $\lim_{\tau\to\infty} e^{-\delta^L(\tau-t)}C(\tilde{\sigma}^L,\tau)$ cannot feasibly exceed $-\infty$. This contradicts the assumption that $\tilde{\sigma}^L$ is a profitable deviation from σ^{L*}.

Open-loop schedule

By (i), the proof that σ^{H*} is a best response to σ^{L*} also serves as a proof that, in the open-loop setting, $\overline{x}^H$ is a best response to $\overline{x}^L$. The proof that, given σ^{H*}, L weakly disprefers reallocations from σ^{L*} (at a node $h_{|t}$) implemented within the $[t,\xi)$ interval also, with $t=0$ and $\xi=\infty$, serves as a proof that, in the open-loop setting, $\overline{x}^L$ is a best response to $\overline{x}^H$.

A.9 Proof of Proposition 9

Existence and uniqueness of constrained-polarized equilibrium schedule

Suppose that, at each post-announcement node $h_{|t}$, the players' spending plans follow

$$\underline{\overline{x}}_s^H = B^H(h_{|t})\overline{\alpha}e^{(r-\overline{\alpha})(s-t)},$$
$$\underline{\overline{x}}_s^L = B^L(h_{|t})\underline{\alpha}e^{(r-\underline{\alpha})(s-t)}.$$

Denote these spending plans $\{\underline{\overline{x}}^i(h_{|t})\}$.

Any feasible deviation by H from to a strategy implementing $x^H \neq \underline{\overline{x}}^H(h_{|t})$, following any node, shifts H's resources from earlier to later times—in the sense of (256)–(258) and the surrounding discussion—without affecting $x^L = \underline{\overline{x}}^L(h_{|t})$. From the $\underline{\overline{X}}(h_{|t})$ baseline, increases in allocations to later times offer lower δ^H-discounted

marginal utility than increases in allocations to earlier times—

$$s_2 > s_1 \implies e^{-(r-\delta^H)(s_2-t)}u'\big(\underline{X}_{s_1}(h_{|t})\big) \leq e^{-(r-\delta^H)(s_1-t)}u'\big(\underline{X}_{s_2}(h_{|t})\big)$$

—since the growth rate of $\underline{X}$ is everywhere a weighted average of $r-\underline{\alpha}$ and $r-\overline{\alpha}$, which weakly exceeds $r-\alpha^H$. The deviation thus lowers H's payoff.

Fixing $x^H = \underline{x}^H(h_{|t})$, any feasible deviation by L likewise only shifts resources from later to earlier times and, since a weighted average of $r-\underline{\alpha}$ and $r-\overline{\alpha}$ is everywhere weakly less $r-\alpha^L$, lowers L's payoff.

Given existence, uniqueness here follows immediately from the definition of constrained polarization.

Open-loop schedule

Since the above applies to any $h_{|t}$, including $h_{|t} = \emptyset$, $\underline{x}$ is also an equilibrium of the open-loop game.

A.10 Proof of Proposition 10

Assume throughout that spending minima and maxima satisfy (125).

Given a spending rate $\underline{\alpha}$ ($\overline{\alpha}$), define $\underline{\delta}$ ($\overline{\delta}$) as the time preference rate for which the corresponding spending rate is optimal, as in (136).

Bounded WTP for H given both constraints

H's payoff given a spending minimum $\underline{\alpha} \geq \alpha^L$ only, $U^H(x[B^H, B^L, \underline{\alpha}, \infty])$, is less than

$$U_{\delta^H}(B^L, \underline{\alpha}) + \left(B^L \underline{\alpha}\right)^{-\gamma} B^H.$$

This is the payoff she receives from L spending B^L at rate $\underline{\alpha}$, as he does under $\underline{x}^L$, plus the payoff-increase implied if she could spend all of B^H at time zero, when the δ^H-discounted marginal utility to resource allocations is highest from the $\underline{x}^L$ baseline, without confronting the diminishing marginal utility to her own spending. Then substituting $B^H(1-w)$ for B^H and using the explicit expression for $U_{\delta^H}(\cdot)$ from (135), we have

$$\begin{aligned} U^H\left(x[B^H(1-w), B^L, \underline{\alpha}, \infty]\right) &< U_{\delta^H}(B^L, \underline{\alpha}) + \left(B^L \underline{\alpha}\right)^{-\gamma} B^H(1-w) \\ &= \frac{\left(B^L\right)^{1-\gamma}}{1-\gamma} \frac{\underline{\alpha}^{1-\gamma}}{\underline{\alpha} + \delta^H - \underline{\delta}} + \left(B^L \underline{\alpha}\right)^{-\gamma} B^H(1-w), \quad \gamma \neq 1; \quad (262) \\ &= \frac{\delta^H \ln\left(B^L \underline{\delta}\right) + r - \underline{\delta}}{\delta^{H2}} + \left(B^L \underline{\delta}\right)^{-1} B^H(1-w), \quad \gamma = 1. \end{aligned}$$

H's payoff given spending minimum $\underline{\alpha} \geq \alpha^L$ and spending maximum $\overline{\alpha} \leq \alpha^H$, $U^H(x[B^H, B^L, \underline{\alpha}, \overline{\alpha}])$, is weakly greater than $U_{\delta^H}(B, \underline{\alpha})$—the payoff she receives when the collective budget B is spent at rate $\underline{\alpha}$—since given $\overline{\underline{x}}^L$, she can achieve this payoff

by choosing $x_t^H = B^H \underline{\alpha} e^{(r-\underline{\alpha})t}$. We thus have

$$\begin{aligned} U^H(x[B^H, B^L, \underline{\alpha}, \overline{\alpha}]) &\geq U_{\delta^H}(B, \underline{\alpha}) \\ &= \frac{B^{1-\gamma}}{1-\gamma} \frac{\underline{\alpha}^{1-\gamma}}{\underline{\alpha} + \delta^H - \underline{\delta}}, \qquad \gamma \neq 1; \\ &= \frac{\delta^H \ln\left(B\underline{\delta}\right) + r - \underline{\delta}}{\delta^{H2}}, \quad \gamma = 1. \end{aligned} \tag{263}$$

Solving for the w^* such that (262)=(263) at $w = w^*$ yields

$$\begin{aligned} w^*(b) &= 1 - \frac{b^\gamma - b}{1-b} \frac{1}{1-\gamma} \frac{\underline{\alpha}}{\underline{\alpha} + \delta^H - \underline{\delta}}, \quad \gamma \neq 1; \\ &= 1 + \frac{b}{1-b} \frac{\ln(b)}{\delta^H}, \qquad\qquad\qquad \gamma = 1. \end{aligned} \tag{264}$$

For the $\gamma \neq 1$ case, by (120) and (111), $\underline{\alpha} \in [\alpha^L, \alpha^H]$ implies that $\underline{\alpha} + \delta^H - \underline{\delta} > 0$. Also, since $b \in (0,1)$, $(b^\gamma - b)/(1-\gamma) > 0$: the numerator and denominator are both positive if $\gamma < 1$ and both negative if $\gamma > 1$. This establishes that $w^*(b) < 1$ for $b \in (0,1)$. Observe also that $w^*(\cdot)$ is continuous in b, with $w^*(0) = 1$ and, by L'Hôpital's Rule,

$$\lim_{b \to 1^-} \frac{b^\gamma - b}{1-b} = 1 - \gamma \quad \Longrightarrow \quad \lim_{b \to 1^-} w^*(b) = 1 - \frac{\underline{\alpha}}{\underline{\alpha} + \delta^H - \underline{\delta}} < 1.$$

For the $\gamma = 1$ case, since $b \in (0,1)$, $\ln(b) < 0$. Also, $\delta^H > 0$ by (111). So $w^*(b) < 1$. Observe again that $w^*(\cdot)$ is continuous in b, with $\lim_{b \to 0^+} w^*(b) = 1$ and $\lim_{b \to 1^-} w^*(b) = 1 - \delta^H < 1$ by L'Hôpital's Rule.

Since increases in H's budget continuously strictly increase $\underline{x}^H$ early in time without affecting $\underline{x}^L$, $U^H\big(x[B^H(1-w), B^L, \underline{\alpha}, \infty]\big)$ continuously strictly decreases in w. Also,

$$U^H\big(x[B^H, B^L, \underline{\alpha}, \infty]\big) \geq U^H\big(x[B^H, B^L, \underline{\alpha}, \overline{\alpha}]\big),$$

since x^L is the same in each case, and

$$U^H\big(x[0, B^L, \underline{\alpha}, \infty]\big) < U^H\big(x[B^H, B^L, \underline{\alpha}, \overline{\alpha}]\big),$$

for the same reason. So, recalling that we can assume $B = 1$ without loss of generality due to homotheticity, there is a unique $w^H(b) \in (0, 1)$ such that

$$U^H\big(x[(1-b)(1-w^H(b)), b, \underline{\alpha}, \infty]\big) = U^H\big(x[(1-b), b, \underline{\alpha}, \overline{\alpha}]\big).$$

Then by (262)–(264),

$$U^H\big(x[(1-b)(1-w^*(b)), b, \underline{\alpha}, \infty]\big) < U^H\big(x[(1-b), b, \underline{\alpha}, \overline{\alpha}]\big).$$

So $w^H(b) < w^*(b) < 1$ for all $b \in (0, 1)$. Since $w^*(b)$ is continuous and $\lim_{b\to 1^-} w^*(b) < 1$, it only remains to show that $\lim_{b\to 0^+} w^H(b) < 1$.

To do this, observe that

$$\begin{aligned} U^H\big(x[B^H(1-w), B^L, \underline{\alpha}, \infty]\big) &\leq U_{\delta^H}\big(B^H(1-w) + B^L, \alpha^H\big) \\ &= \frac{\big(B^H(1-w) + B^L\big)^{1-\gamma}}{1-\gamma}\big(\alpha^H\big)^{-\gamma}, \qquad \gamma \neq 1; \qquad (265) \\ &= \frac{\delta^H \ln\big((B^H(1-w) + B^L)\delta^H\big) + r - \delta^H}{\delta^{H2}}, \quad \gamma = 1, \end{aligned}$$

with the first weak inequality following from the fact that the right-hand side is the maximum utility achievable for H with collective budget $B^H(1-w) + B^L$. Solving

for the $\tilde{w}$ such that (265)=(263) at $w = \tilde{w}$ yields

$$\tilde{w}(b) = \frac{1}{1-b}\,\frac{\alpha^H\eta - \underline{\alpha}}{\alpha^H\eta}, \tag{266}$$

where η equals (110) with $\alpha^H, \delta^H, \overline{\alpha}, \overline{\delta}$ in place of $\alpha, \delta, \tilde{\alpha}, \tilde{\delta}$.

As above, we must have $w^H(b) < \tilde{w}(b) < 1$ for all $b \in (0,1)$. Since $\tilde{w}(b)$ is continuous and $\tilde{w}(0) < 1$, $\lim_{b\to 0^+} w^H(b) < 1$, as desired. So $\sup_b w^H(b, \underline{\alpha}, \overline{\alpha}) < 1$ given $\underline{\alpha} \geq \alpha^L$ and $\underline{\alpha} \leq \alpha^H$.

Bounded WTP for H given spending maximum only

Given $B = 1$, H's payoff in the absence of constraints, after foregoing fraction w of her budget, can be found analytically:

$$\begin{aligned} U^H\big(x^*[(1-b)(1-w), b]\big) &= \int_0^\infty e^{-\delta^H t} u\big(x_t^*[(1-b)(1-w), b]\big)dt \\ &= \frac{b^{1-\gamma}}{1-\gamma}\big(\alpha^L\big)^{-\gamma}\Big(1 + \frac{(1-b)(1-w)\alpha^H}{b\alpha^L}\eta\Big)^{-\gamma}\Big(\frac{1-b}{b}\eta^\gamma + \frac{\alpha^L}{\alpha^L + \delta^H - \delta^L}\Big), \quad \gamma \neq 1; \\ &= \frac{1}{\delta^H}\Big(\ln\Big(b\delta^L\Big(1 + \frac{(1-b)(1-w)\alpha^H}{b\alpha^L}\eta\Big)\Big) + \frac{r-\delta^L}{\delta^H}\Big), \qquad \gamma = 1, \end{aligned} \tag{267}$$

where η equals (110) with α^H, δ^H, α^L, δ^L in place of α, δ, $\tilde{\alpha}$, $\tilde{\delta}$.

H's payoff given spending maximum $\overline{\alpha} \leq \alpha^H$ only is weakly greater than the payoff she receives when the collective budget $B = 1$ is spent at rate α^L (with equality only when b is sufficiently large: see (124)):

$$U^H(x[1-b, b, 0, \overline{\alpha}]) \geq U_{\delta^H}(1, \alpha^L). \tag{268}$$

Solving for the w^* such that (267)=$U_{\delta^H}(1, \alpha^L)$ at $w = w^*$ yields

$$w^*(b) = 1 - \frac{\alpha^L}{\alpha^H \eta} < 1, \tag{269}$$

independent of b.

$U^H\big(x^*[(1-b)(1-w), b]\big)$ strictly and continuously decreases in w, since $\partial U^H(x^*[B^H, B^L])/\partial B^H > 0$ everywhere (see (220)).

$$U^H(x^*[1-b, b]) > U^H\big(x[1-b, b, 0, \overline{\alpha}]\big),$$

by Proposition 4 and the fact that $\overline{x}^L$ is a best response to $\overline{x}^H$ even in the absence of constraints (see Appendix A.7), so that there would be alternative equilibrium of the Stackelberg game (with $x^H = \overline{x}^H$) if the inequality above failed. Finally,

$$U^H(x^*[0, b]) < U^H\big(x[1-b, b, 0, \overline{\alpha}]\big),$$

since $x^*[0, b] = x[0, b, 0, \overline{\alpha}]$ is the δ^L-optimal schedule given collective budget b, and $x_t[B^H, b, 0, \overline{\alpha}]$ strictly increases in B^H for all t. So there is a unique $w^H(b) \in (0, 1)$ such that

$$U^H(x^*[(1-b)(1-w^H(b)), b]) = U^H\big(x[1-b, b, 0, \overline{\alpha}]\big).$$

Then by (268)–(269),

$$U^H(x^*[(1-b)(1-w^*(b)), b]) \leq U^H\big(x[1-b, b, 0, \overline{\alpha}]\big).$$

So $w^H(b) \leq w^*(b) < 1$ for all $b \in (0, 1)$, and $w^H(b)$ is bounded below 1 as b approaches 0 or 1. So $\sup_b w^H(b, 0, \overline{\alpha}) < 1$ given $\overline{\alpha} \leq \alpha^H$.

Unbounded WTP for L given both constraints

Assume that $b < 1 - \alpha^L/\overline{\alpha}$, so that $\overline{t}^* \equiv \overline{t}^*[1-b, b(1-w)] > 0$ for all $w \in [0,1)$ (see (123) and the discussion around (124)).

Given $B = 1$ and a spending maximum $\overline{\alpha} \leq \alpha^H$ only, L's payoff after foregoing fraction w of his budget can be found analytically:

$$U^L\big(x[(1-b), b(1-w), 0, \overline{\alpha}]\big) = \int_0^{\overline{t}^*} e^{-\delta^L t} u\big((1-b)\overline{\alpha} e^{(r-\overline{\alpha})t}\big) dt + \int_{\overline{t}^*}^{\infty} e^{-\delta^L t} u\Big(\big((1-b)e^{(r-\overline{\alpha})\overline{t}^*} + b(1-w)e^{r\overline{t}^*}\big) e^{(r-\alpha^L)(t-\overline{t}^*)}\Big) dt$$

$$= \frac{\big((1-b)\overline{\alpha}\big)^{1-\gamma}}{1-\gamma}\Big(\frac{1}{\overline{\alpha}+\delta^L-\overline{\delta}} + \Big(\frac{1}{\alpha^L} - \frac{1}{\overline{\alpha}+\delta^L-\overline{\delta}}\Big)\Big(\frac{1-b}{b(1-w)}\,\frac{\overline{\alpha}-\alpha^L}{\alpha^L}\Big)^{-\frac{\overline{\alpha}+\delta^L-\overline{\delta}}{\overline{\alpha}}}\Big), \ \gamma \neq 1; \tag{270}$$

$$= \frac{1}{\delta^L}\Big(\ln\big((1-b)\overline{\delta}\big) + \frac{r-\overline{\delta}}{\delta^L} + \Big(\frac{1-b}{b(1-w)}\Big)^{-\frac{\delta^L}{\overline{\delta}}}\Big(\frac{\overline{\delta}-\delta^L}{\delta^L}\Big)^{\frac{\overline{\delta}-\delta^L}{\overline{\delta}}}\Big), \qquad \gamma = 1.$$

Observe that, by (111) and (120), $\overline{\alpha} + \delta^L - \overline{\delta} > 0$.

Let $\tilde{t}(b) \equiv -\ln(b)z$, where

$$z \equiv \begin{cases} \frac{\overline{\delta}-\delta^L}{2\overline{\alpha}(\overline{\delta}-\underline{\alpha}-\delta^L)}, & \underline{\alpha} < \overline{\delta} - \delta^L; \\ 1, & \underline{\alpha} = \overline{\delta} - \delta^L; \\ \infty, & \underline{\alpha} > \overline{\delta} - \delta^L. \end{cases} \tag{271}$$

L's payoff given both spending minimum $\underline{\alpha} \geq \alpha^L$ and spending maximum $\overline{\alpha} \in (\alpha^L, \alpha^H]$ can be upper-bounded by the sum of the following three terms. First, as a

baseline, is the payoff he receives from H's schedule $\underline{\overline{x}}^H$ in isolation:

$$\begin{aligned} U_{\delta^L}(1-b,\overline{\alpha}) &= \frac{(1-b)^{1-\gamma}}{1-\gamma}\,\frac{\overline{\alpha}^{1-\gamma}}{\overline{\alpha}+\delta^L-\overline{\delta}}, \qquad \gamma \neq 1; \\ &= \frac{\delta^L \ln\big((1-b)\overline{\delta}+r-\overline{\delta}\big)}{\delta^{L2}}, \quad \gamma = 1. \end{aligned} \tag{272}$$

Second is the additional payoff he would receive from his own spending from $t=0$ to $\tilde{t}(b)$ if this spending confronted no diminishing marginal utility, so that each unit spent at t offered $u'(\underline{\overline{x}}_t^H)$ units of flow utility:

$$\begin{aligned} &\int_0^{\tilde{t}(b)} e^{-\delta^L t} b\underline{\alpha} e^{(r-\underline{\alpha})t}\big((1-b)\overline{\alpha}e^{(r-\overline{\alpha})t}\big)^{-\gamma} dt \\ &= b\underline{\alpha}\big((1-b)\overline{\alpha}\big)^{-\gamma}\frac{1}{\underline{\alpha}+\delta^L-\overline{\delta}}\Big(1-b^{-(\overline{\delta}-\underline{\alpha}-\delta^L)z}\Big). \end{aligned} \tag{273}$$

Third is the additional payoff he would receive from the present value of his budget not spent before $\tilde{t}(b)$, i.e. $be^{-\underline{\alpha}\tilde{t}(b)}$, if it were feasible for him to spend it subject to no constraints and subject to no diminishing marginal utility from L's own other spending. This can be found by subtracting $U_{\delta^L}(1-b,\overline{\alpha})$ from (270) and substituting $e^{-\underline{\alpha}\tilde{t}(b)}$ for $1-w$ in the remainder to get, for all γ,

$$\frac{\overline{\alpha}^{1-\gamma}}{\overline{\alpha}+\delta^L-\overline{\delta}}\Big(\frac{\overline{\alpha}-\alpha^L}{\alpha^L}\Big)^{\frac{\overline{\delta}-\delta^L}{\overline{\alpha}}}(1-b)^{\frac{\overline{\delta}-\delta^L-\gamma\overline{\alpha}}{\overline{\alpha}}}b^{(1+\underline{\alpha}z)\frac{\overline{\alpha}+\delta^L-\overline{\delta}}{\overline{\alpha}}}. \tag{274}$$

Denote the sum of (272)–(274) by $\tilde{U}^L(b)$, so that

$$U^L\big(x[1-b,b,\underline{\alpha},\overline{\alpha}]\big) \leq \tilde{U}^L(b), \tag{275}$$

and let $w^*(b)$ denote the value of w such that (270)=$\tilde{U}^L(b)$ at $w = w^*(b)$:

$$w^*(b) = 1 - \Bigg((\overline{\alpha} + \delta^L - \overline{\delta}) \Big(\frac{\alpha^L}{\overline{\alpha} - \alpha^L} \Big)^{\frac{\overline{\delta} - \delta^L}{\overline{\alpha}}} (1-b)^{\frac{\overline{\alpha} + \delta^L - \overline{\delta}}{\overline{\alpha}}} \tag{276}$$
$$\Big(\frac{\underline{\alpha} \big((1-b)\overline{\alpha} \big)^{-\gamma}}{\underline{\alpha} + \delta^L - \overline{\delta}} \big(b^{\frac{\overline{\delta} - \delta^L}{\overline{\alpha}}} - b^{\frac{\overline{\delta} - \delta^L}{\overline{\alpha}} - (\overline{\delta} - \underline{\alpha} - \delta^L)z} \big)$$
$$+ \frac{\overline{\alpha}^{1-\gamma}}{\overline{\alpha} + \delta^L - \overline{\delta}} \big(\frac{\overline{\alpha} - \alpha^L}{\alpha^L} \big)^{\frac{\overline{\delta} - \delta^L}{\overline{\alpha}}} (1-b)^{\frac{\overline{\delta} - \delta^L - \gamma\overline{\alpha}}{\overline{\alpha}}} b^{(1 + \underline{\alpha} z) \frac{\overline{\alpha} + \delta^L - \overline{\delta}}{\overline{\alpha}}} \Big) \Bigg)^{\frac{\alpha}{\overline{\alpha} + \delta^L - \overline{\delta}}}$$

$\lim_{b \to 0} w^*(b) = 1$ under all conditions.

Recall that $w^L(b, \underline{\alpha}, \overline{\alpha})$ is the value of w such that

$$U^L\big(x[1-b, b(1-w), 0, \overline{\alpha}]\big) = U^L\big(x[1-b, b, \underline{\alpha}, \overline{\alpha}]\big).$$

$w^L(b, \underline{\alpha}, \overline{\alpha}) \in (0,1)$ is defined by the facts that

$$U^L\big(x[1-b, b, 0, \overline{\alpha}]\big) > U^L\big(x[1-b, b, \underline{\alpha}, \overline{\alpha}]\big)$$

(since L can implement $x[1-b, b, \underline{\alpha}, \overline{\alpha}]$ in the presence of a spending maximum but no minimum, in the open-loop setting, but prefers not to (see Appendix A.7)); that

$$U^L\big(x[1-b, 0, 0, \overline{\alpha}]\big) < U^L\big(x[1-b, b, \underline{\alpha}, \overline{\alpha}]\big);$$

and that $U^L\big(x[1-b, b(1-w), 0, \overline{\alpha}]\big)$ strictly and continuously decreases in w.

Furthermore by (275) $w^L(b, \underline{\alpha}, \overline{\alpha}) \geq w^*(b)$ for all b. So $\lim_{b \to 0} w^L(b, \underline{\alpha}, \overline{\alpha}) = 1$.

Unbounded WTP for L given spending minimum only

Given $B = 1$, L's payoff in the absence of constraints, after foregoing fraction w of his budget, can be found analytically:

$$
\begin{aligned}
&U^L\big(x^*[1-b, b(1-w)]\big) = \int_0^\infty e^{-\delta^L t} u\big(x_t^*[1-b, b(1-w)]\big)dt \\
&= \frac{1}{1-\gamma}\Bigg[\frac{\Big((1-b)\alpha^H + b(1-w)\alpha^L/\eta\Big)^{1-\gamma}}{\alpha^H + \delta^L - \delta^H}\Bigg(1 - \Big(b(1-w) + \frac{(1-b)\alpha^H}{\alpha^L}\eta\Big)^{-\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}} \\
&\quad \cdot \big(b(1-w)\big)^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}}\Bigg) + (\alpha^L)^{-\gamma}\Big(b(1-w) + \frac{(1-b)\alpha^H}{\alpha^L}\eta\Big)^{-\gamma\frac{\alpha^L}{\alpha^H}}\big(b(1-w)\big)^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}}\Bigg], \quad \gamma \neq 1; \\
&= \frac{1}{\delta^L}\Bigg[\ln\Big((1-b)\delta^H + \frac{b(1-w)\delta^L}{\eta}\Big) + \frac{r-\delta^H}{\delta^L} \qquad (277) \\
&\quad + \frac{(\delta^H-\delta^L)^2}{\delta^H\delta^L}\Big(b(1-w) + \frac{(1-b)\delta^H}{\delta^L}\eta\Big)^{-\frac{\delta^L}{\delta^H}}\big(b(1-w)\big)^{\frac{\delta^L}{\delta^H}}\Bigg], \quad \gamma = 1,
\end{aligned}
$$

where η equals (110) with α^H, δ^H, α^L, δ^L in place of α, δ, $\tilde{\alpha}$, $\tilde{\delta}$.

In the presence of a spending minimum $\underline{\alpha}$ only, the collective spending rate at time 0—$\underline{X}_0$—is weakly less than α^H, the spending rate at time 0 that H implements if the entire collective budget is hers. (If it were greater, then because $\underline{X}$ grows at rate $r - \alpha^H$ until $\underline{t}^* \leq \infty$ and then grows at $r - \underline{\alpha} \geq r - \alpha^H$, $\underline{X}$ would be unaffordable.) Thus we also have $\tilde{t}[b] \geq \underline{t}^*[b]$, where

$$\tilde{t}[b] \equiv \ln\Big(\frac{\alpha^H}{b\underline{\alpha}}\Big)/(\alpha^H - \underline{\alpha})$$

denotes the regime-change time t at which $\alpha^H e^{(r-\alpha^H)t} = b\underline{\alpha}e^{(r-\underline{\alpha})t}$.

So given spending minimum $\underline{\alpha}$ only, L's payoff is weakly less than $\tilde{U}^L(b)$, where $\tilde{U}^L(b)$ denotes the payoff he receives if $X_t = \alpha^H e^{(r-\alpha^H)t}$ for $t \in [0, \tilde{t}[b])$ and $= b\underline{\alpha}e^{(r-\underline{\alpha})t}$

for $t \in [\tilde{t}[b], \infty)$.

We will assume in the expression below that $\underline{\alpha} < \alpha^H$. We can ignore the $\underline{\alpha} = \alpha^H$ case because, for any $b > 0$,

$$U^L\big(x[1-b, b, \alpha^H, \infty]\big) < U^L\big(x[1-b, b, \underline{\alpha}, \infty]\big) \quad \forall \underline{\alpha} < \alpha^H :$$

this follows from the fact that in the constrained open-loop setting with $\underline{\alpha} < \alpha^H$ it is feasible for L, given $x^H = \underline{x}^H$, to implement the δ^H-optimal allocation of the collective budget, but he prefers not to do so (see Appendix A.8). We can thus upper-bound $U^L\big(x[1-b, b, \alpha^H, \infty]\big)$ by the expression below for any choice of $\underline{\alpha} \in (\alpha^L, \alpha^H)$.

$$U^L\big(x[1-b, b, \underline{\alpha}, \infty]\big) \leq \tilde{U}^L(b) \tag{278}$$

$$= \frac{1}{1-\gamma}\Big(\frac{\big(\alpha^H\big)^{1-\gamma}}{\alpha^H + \delta^L - \delta^H}\Big(1 - \Big(\frac{b\underline{\alpha}}{\alpha^H}\Big)^{\frac{\alpha^H + \delta^L - \delta^H}{\alpha^H - \underline{\alpha}}}\Big) + \frac{\big(\alpha^H\big)^{1-\gamma}}{\underline{\alpha} + \delta^L - \underline{\delta}}\Big(\frac{b\underline{\alpha}}{\alpha^H}\Big)^{\frac{\alpha^H + \delta^L - \delta^H}{\alpha^H - \underline{\alpha}}}\Big), \quad \gamma \neq 1; \tag{279}$$

$$= \frac{1}{\delta^L}\Big(\ln\big(\delta^H\big) + \frac{\delta^H - \underline{\delta}}{\delta^L}\Big(\frac{b\underline{\delta}}{\delta^H}\Big)^{\frac{\delta^L}{\delta^H - \underline{\delta}}} + \frac{r - \delta^H}{\delta^L}\Big), \quad \gamma = 1.$$

Let $w^*(b)$ denote the value of $w \in (-\infty, 1)$ such that (277)=(279) at $w = w^*(b)$. $w^*(b)$ is defined, since $U^L(x^*[1-b, b(1-w)])$ strictly and continuously decreases in w from the supremum feasible payoff at $w = -\infty$ (∞ if $\gamma \leq 1$, 0 if $\gamma > 1$) to $U_{\delta^L}(1-b, \delta^H)$ at $w = 1$, and (279) is greater than $U_{\delta^L}(1-b, \alpha^H)$. Furthermore, there cannot be a sequence $b_n \to 0$ with $\lim_{n\to\infty} b_n(1 - w^*(b_n)) = \underline{b} > 0$, because by the continuity of $U^L(x^*[B^H, B^L])$ in both budgets and the continuity of $\tilde{U}^L(b)$, this would imply

$$U^L\big(x^*[1, \underline{b}]\big) = \tilde{U}^L(0),$$

but $\tilde{U}^L(0) = U^L(x^*[1, 0]) = U_{\delta^L}(1, \alpha^H)$, and $U^L(x^*[B^H, B^L])$ strictly increases in B^L

(see Appendix A.5: Final period: Unique equilibrium schedule given placement of first announcement). So $\lim_{n\to\infty} b_n(1 - w^*(b_n)) = 0$.

To find $w^*(b)$, set (277) = (279) and subtract $U_{\delta^L}(1, \alpha^H)$ from both sides. If $\underline{\alpha} \leq \delta^H - \delta^L$, divide both sides by $b^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H-\underline{\alpha}}}$ and simplify to get

$$\begin{aligned}
&\frac{1}{\alpha^H+\delta^L-\delta^H}\Bigg(\Big((1-b)\alpha^H + b\big(1-w^*(b)\big)\frac{\alpha^L}{\eta}\Big)^{1-\gamma} - \alpha^H\Bigg)b^{-\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H-\underline{\alpha}}} \\
&\quad + \big(\alpha^L\big)^{-\gamma}\Big(1 - \frac{\alpha^L}{\alpha^L+\delta^H-\delta^L}\,\frac{\alpha^H}{\alpha^H+\delta^L-\delta^H}\Big)b^{-\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}\,\frac{\underline{\alpha}}{\alpha^H-\underline{\alpha}}}\big(1-w^*(b)\big)^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H}} \\
&\quad = \Big(\frac{1}{\underline{\alpha}+\delta^L-\underline{\delta}} - \frac{1}{\alpha^H+\delta^L-\delta^H}\Big)(\alpha^H)^{-\frac{\underline{\alpha}+\delta^L-\underline{\delta}}{\alpha^H-\underline{\alpha}}}, \qquad \gamma \neq 1; \\
&\Bigg(\ln\Big((1-b)\delta^H + \frac{b\big(1-w^*(b)\big)\delta^L}{\eta}\Big) - \ln(\delta^H)\Bigg)b^{-\frac{\delta^L}{\delta^H-\underline{\delta}}} \\
&\quad + \frac{(\delta^H-\delta^L)^2}{\delta^H\delta^L}\Big(b\big(1-w^*(b)\big) + \frac{(1-b)\delta^H}{\delta^L}\eta\Big)^{-\frac{\delta^L}{\delta^H}}b^{-\frac{\delta^L}{\delta^H}\,\frac{\underline{\delta}}{\delta^H-\underline{\delta}}}\big(1-w^*(b)\big)^{\frac{\delta^L}{\delta^H}} \\
&\quad = (\delta^H-\underline{\delta})\Big(\frac{\underline{\delta}}{\delta^H}\Big)^{\frac{\delta^L}{\delta^H-\underline{\delta}}}, \qquad \gamma = 1.
\end{aligned}$$

In each case, given $\underline{\alpha} \leq \delta^H - \delta^L$, the exponent on b in the first term is greater than -1. So the limit of the first term as $b \to 0$ is finite, regardless of $w^*(\cdot)$, by L'Hôpital's Rule. The coefficient to the left of the b in the second term is nonzero by (120). So, from the second term, unless $\lim_{b\to 0} w^*(b) = 1$, the limit of left-hand side as $b \to 0$ is not a finite constant, as the right-hand side is.

If $\underline{\alpha} > \delta^H - \delta^L$, instead divide both sides by b—i.e. multiply the expression above by $b^{\frac{\alpha^H+\delta^L-\delta^H}{\alpha^H-\underline{\alpha}}-1}$. Note that if $\underline{\alpha} > \delta^H - \delta^L$, the exponent on b on the right-hand side is now positive, so the limit of the right-hand side as $b \to 0$ is zero. Here the limit of the first term on the left-hand side as $b \to 0$ is finite, regardless of $w^*(\cdot)$, by L'Hôpital's Rule, and limit of the second term is defined and finite only if $\lim_{b\to 0} w^*(b) = 1$.

So $\lim_{b\to 0} w^*(b) = 1$.

Recall that $w^L(b, \underline{\alpha}, \infty)$ is the value of w such that

$$U^L\big(x^*[1-b, b(1-w)]\big) = U^L\big(x[1-b, b, \underline{\alpha}, \infty]\big).$$

$w^L(b)$ is defined and less than 1 by the reasoning following (279), noting that $U^L\big(x[1-b, b, \underline{\alpha}, \infty]\big)$ too is greater than $U_{\delta^L}(1-b, \alpha^H)$. Then by (278) and the fact that $U^L(x^*[1-b, b(1-w)])$ decreases in w, $w^L(b, \underline{\alpha}, \infty) \geq w^*(b)$ for all b.

So $\lim_{b\to 0} w^L(b, \underline{\alpha}, \infty) = 1$.

A.11 Proof of Proposition 11

By (237) and the surrounding discussion, a collective schedule is efficient iff it is a measure-zero deviation from a collective schedule X that exhausts the collective budget and satisfies

$$X_t = X_0 e^{\frac{r}{\gamma}t}\big(ae^{-\delta^H t} + (1-a)e^{-\delta^L t}\big)^\gamma$$

for some X_0 and some $a \in [0, 1]$ representing the weight placed on U^H (with weight $1-a$ placed on U^L). The result follows from comparing this expression to the expressions for $x[\underline{\alpha}, \overline{\alpha}, B^H, B^L]$ from Propositions 5 and 7–9.

A.12 Proof of Proposition 12

Suppose $\{u(\cdot)\}$ satisfies single crossing from X, and let $\underline{s} < \overline{s}$.

$$\frac{\partial U^H(X)}{\partial X_{\overline{s}}} \geq \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^H(X)}{\partial X_{\underline{s}}} \tag{280}$$

$$\begin{aligned} \implies \exists \tilde{t} : & \sum_{t=0}^{\tilde{t}-1} e^{-\sum_{q=1}^{t} \delta_q^H} \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \\ & + \sum_{t=\tilde{t}}^{\infty} e^{-\sum_{q=1}^{t} \delta_q^H} \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \geq 0 \end{aligned} \tag{281}$$

with each term in the first sum non-positive and each term in the second sum non-negative, by (127), and at least one term in the second sum positive, by (128).

If (281) holds for $\tilde{t} = 0$, then we immediately have

$$\begin{aligned} \sum_{t=\tilde{t}}^{\infty} e^{-\sum_{q=1}^{t} \delta_q^L} \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) > 0 & \\ \implies \frac{\partial U^L(X)}{\partial X_{\overline{s}}} > \frac{R_{\underline{s}}}{R_{\overline{s}}} \frac{\partial U^L(X)}{\partial X_{\underline{s}}}. & \end{aligned} \tag{282}$$

Otherwise, since $e^{\sum_{q=1}^{t} (\delta_q^H - \delta_q^L)} - 1$ equals 0 at $t = 0$ and strictly increases in t, (281) implies

$$\begin{aligned} & \sum_{t=0}^{\tilde{t}-1} \Big(e^{-\sum_{q=1}^{t} \delta_q^L} - e^{-\sum_{q=1}^{t} \delta_q^H} \Big) \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \\ & + \sum_{t=\tilde{t}}^{\infty} \Big(e^{-\sum_{q=1}^{t} \delta_q^L} - e^{-\sum_{q=1}^{t} \delta_q^H} \Big) \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \geq \end{aligned} \tag{283}$$

$$\begin{aligned} & \sum_{t=0}^{\tilde{t}-1} \Big(e^{\sum_{q=1}^{\tilde{t}} (\delta_q^H - \delta_q^L)} - 1 \Big) e^{-\sum_{q=1}^{t} \delta_q^H} \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \\ & + \sum_{t=\tilde{t}}^{\infty} \Big(e^{\sum_{q=1}^{\tilde{t}} (\delta_q^H - \delta_q^L)} - 1 \Big) e^{-\sum_{q=1}^{t} \delta_q^H} \Big(\frac{\partial u_t(X)}{\partial X_{\overline{s}}} - \frac{R_{\overline{s}}}{R_{\underline{s}}} \frac{\partial u_t(X)}{\partial X_{\underline{s}}} \Big) \geq 0, \end{aligned} \tag{284}$$

with inequality (283) holding strictly unless the difference in derivatives equals zero for all $t \neq \tilde{t}$, in which case inequality (284) holds strictly by (128).

So the left-hand side of (283) is positive. In conjunction with (281), this implies (282). This proves part (a).

In the generalized open-loop game, suppose the players choose a strategy profile x such that $x^H_{\overline{s}} > 0, x^L_{\underline{s}} > 0$ for $\overline{s} > \underline{s}$. Then either (280) holds, or H can increase her payoff by shifting her resources marginally from $\overline{s}$ to $\underline{s}$. If (280) holds, then by part (a), (282) holds, so L can increase his payoff by shifting his resources marginally from $\underline{s}$ to $\overline{s}$. In either case, x is not an equilibrium. This proves part (b).

III Ethical Consumerism

Abstract

I study a static production economy in which consumers have not only preferences over their own consumption but also external, or "ethical", preferences over the supply of each good. Though existing work on the implications of external preferences assumes price-taking, I show that ethical consumers generically prefer not to act even approximately as price-takers. I therefore introduce a near-Nash equilibrium concept that generalizes the near-Nash equilibria found in literature on strategic foundations of general equilibrium to accommodate ethical preferences. I find (narrow) sufficient criteria under which such equilibria exist, and characterize consumer behavior in all such equilibria. Finally I find that ethical preferences can have arbitrary impacts on consumer behavior in equilibrium, including motivating a consumer (1) to decrease her consumption of all goods which she would prefer in greater supply and vice-versa, or (2) not to exhaust her budget, even if her utility increases both in her consumption and in the supply of all goods.

I owe thanks especially to Kevin Roberts and Alex Teytelboym for helpful advice on this paper.

1 Introduction

1.1 Motivation

A individual's spending behavior affects not only the quantities of each good that she herself consumes, but also the total quantities of each good supplied. She may have preferences over all these quantities, and she may optimize her purchasing behavior accordingly. This paper explores a model of general equilibrium in which individuals do so.

Consider, for example, a consumer with concern for animal welfare. The consumer's utility is increasing in his own meat consumption, holding supply fixed, but decreasing in total meat supply: in the supply of pigs and chickens in particular. The latter effect may, and often does, motivate consumers to purchase less meat than they would otherwise (or none at all).

It bears emphasizing that, despite the well-known result (Roberts and Postlewaite, 1976) that a consumer's ability to impact equilibrium prices generically vanishes as the economy grows large, her ability to impact equilibrium supply does not. This is because, as the economy grows large and the price impacts of an individual consumer's demand behavior shrink to zero, any such price impact influences the purchases of a number of other consumers that rises to infinity.

Consider a consumer's decision to buy one more unit of some good at any given price. Compare (a) the impact of this decision on that good's equilibrium prices and supply levels in a given economy with (b) the impact of the decision on equilibrium prices and supply levels in a "doubled" economy with twice as many agents but an identical distribution of endowments, preferences, profit shares, and production technologies. In the doubled economy, quantities supplied and demanded at any

given price will double:

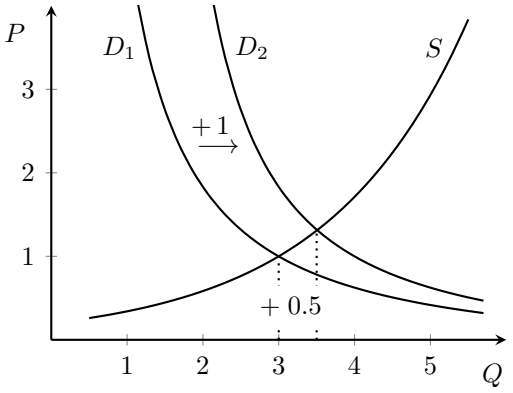


Fig. 1a: Impact on equilibrium quantity of a unit demand shock, original economy

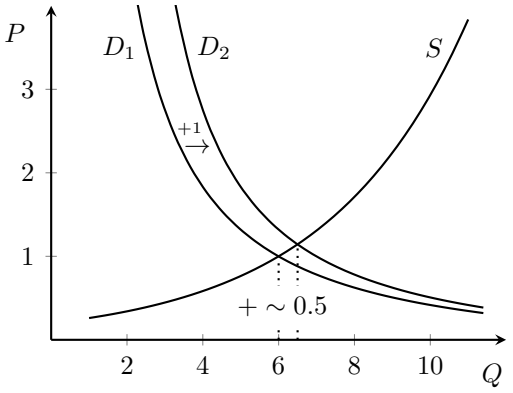


Fig. 1b: Impact on equilibrium quantity of a unit demand shock, doubled economy

In the doubled economy, the units on the vertical axis do not change, but the units on the horizontal axis are doubled, as the rightward movement of the demand curve on the page resulting from a one-unit decrease in demand is halved. Thus, if buying one unit of some good causes its production to increase by 0.5 in an economy with, say, one billion participants, it also causes its production to increase by approximately 0.5 in the economy with two billion participants. The size of an "ethical externality" in this sense does not in general fall to zero as price impacts fall to zero and an economy approaches perfect competition.[1,2]

In light of these ethical externalities, how should an animal-welfare-conscious consumer adjust his demands, relative to what they would be if he had no concern for

[1]Kaufmann et al. (2024) offer survey evidence that many consumers believe that their consumption decisions affect supply (either one-for-one or, as in Figure 1, partially but significantly), and summarize the literature that many consumers do in fact have ethical preferences over the aggregate supply of some goods strong enough to motivate non-negligible shifts in consumption.

[2]Note that this insight comes out only when we define competitive behavior in the usual way, as the limit of behavior across economies with finite but ever larger populations. Under the Aumann (1964) approach of modeling the population as a continuum, demand choices by individuals do not affect prices or supply.

animal welfare? If he believes that the production of a dollar's worth of chicken creates more misery than the production of a dollar's worth of pork, he may naively be inclined to prioritize reducing his purchases of chicken over reducing his purchases of pork. If the supply of pork is more price-elastic and demand for pork less price-elastic than that of chicken, however, this inclination may be misguided. Buying less chicken in this case simply causes the price to fall and the quantity demanded by other consumers to rise, with little net impact on the quantity of chicken consumed. Buying less pork, by contrast, generates a substantial decrease to the quantity of pork consumed. Cutting back on pork may thus be the higher priority.

Complicating matters further, our consumer must consider the impact of his purchases of a good not only on the quantity of that good, but on the quantities of all the goods he cares about. If buying less chicken causes other consumers to substitute to chicken from other meat products, whereas buying less pork causes other consumers to substitute to pork from vegetables, then cutting back on chicken may be the best policy after all.

Though this paper is intended primarily as a model of ethical consumerism, and the precise modeling assumptions made will be tailored to the ethical consumerist context, other agents with preferences over total supply levels also face the motivations and challenges described above. Suppose, for instance, that some goods impose more conventional (i.e. not "ethical") externalities on a consumer, and that these externalities depend on the goods' absolute supply levels rather than on supply per person. Then the utility-impacts of the consumer's contributions to supply are not in general close to zero even when her proportional contributions are small. When she decides what to buy, she too must consider the impacts of her purchases on the equilibrium supply levels of all the goods that impose externalities on her.

This paper aims to characterize, in light of these complications, equilibrium market behavior by "ethical consumers"—and other consumers with preferences over own good-consumption levels and total supply levels—in a competitive production economy. That is, we will study strategic consumer behavior in general equilibrium with externalities.

1.2 Related literature

Though this appears to be the first paper to incorporate all three features (strategic behavior, general equilibrium, and externalities), there exist literatures exploring all three pairs of these features in isolation.

First, there is an extensive literature exploring strategic behavior in general equilibrium without externalities. A central concern of this literature is to put the standard Walrasian model—in which agents are forbidden from strategic behavior and must act as price-takers—on a strategic footing. This is done by constructing a game in which firms and/or individuals can choose their supply or demand correspondences, respectively, and equilibrium prices set the excess demands implied by these chosen correspondences equal to zero. Consider a sequence of such games set in ever larger "replicated economies", with each agent becoming an ever smaller part of the whole, and consider the sequence of Nash equilibria of these games. We can then ask under what circumstances, and in what sense, these Nash equilibria—these profiles of chosen supply and demand correspondences—can or must converge to the "Walrasian" profile in which agents all choose their price-taking supply or demand correspondences.

Roberts and Postlewaite (1976) show that, in an exchange economy, the demand

correspondences chosen in a sequence of Nash equilibria do not generally converge, even pointwise, to the Walrasian demand correspondences. They do find, however, that a consumer's impacts on equilibrium prices, equilibrium allocations, and her own utility by choosing her Walrasian demand correspondence fall to zero as the economy is replicated. It follows that if a consumer faces arbitrarily small costs—e.g. computational costs—to deviating from price-taking behavior, she will choose to act as a price-taker in a sufficiently large economy.

Otani and Sicilian (1990), studying a restriction to the same game, demonstrate that if consumers can only choose differentiable demand functions, there are sequences of Nash equilibria that, again, converge to Walrasian equilibria. Jackson and Manelli (1997) find that uncertainty about others' chosen strategies can also motivate consumers to adopt behavior that converges to fully price-taking behavior.

Though these papers all discuss exchange economies, they are perhaps the papers on strategic foundations for general equilibrium most relevant to this one, which will feature a production economy but strategic behavior only by consumers. Nevertheless, the literature contains numerous other convergence results, including results set in production economies featuring strategic behavior by producers. The message generally taken from this literature appears to be that Walrasian equilibrium reasonably approximates what one should expect to obtain in an economy in which prices result from the strategic behavior of consumers and/or firms. As we will see, and as suggested by the informal discussion of Section 1.1, the presence of consumers with ethical preferences, or supply externalities more generally, allows no such approximation.

A second literature explores general equilibrium with externalities in settings in which individuals are not permitted to act strategically.

Dufwenberg et al. (2011), for instance, present a model intended to capture equilibrium purchasing behavior in a production economy in which individuals may have "other-regarding preferences": preferences defined not only over their own consumption baskets but also over the consumption baskets, utility levels, or budget sets achieved by other consumers. If Dufwenberg et al. had allowed consumers to choose their demand functions strategically, their project would have encompassed the project of the present paper, since concern for the total production of various goods is a special case of concern for others' consumption baskets. Instead, however, Dufwenberg et al. maintain the classical assumption that each individual is *fully* price-taking, in the sense that she acts as if her own purchases have no impact on prices and thus no impact on aggregate production. They thus conclude that an individual's other-regarding preferences have no impact on her purchasing behavior unless they are not separable from her self-regarding preferences: for example, if she prefers brown bread to white bread if and only if her neighbor has two cars.

A similar "fully price-taking" approach to other-regarding preferences, or preferences over total production levels, is taken by Kreps (1990) (p. 203), Ellickson (1994) (Section 7.3), Sobel (2009), and others.

The approach taken in this literature may be motivated by the observation from the literature on strategic foundations for general equilibrium that, in large economies and in the absence of externalities, relatively mild assumptions can guarantee that strategic behavior differs little (or produces outcomes that differ little) from price-taking behavior. In contrast, the present paper is centered on the insight that, as noted above, externalities such as those which may be induced by ethical preferences will typically motivate strategic consumer behavior that differs substantially from price-taking behavior, however large the economy relative to each consumer.

Finally, three strands of literature explore the implications of strategic demand behavior by individuals with preferences over supply levels, but do so in a partial equilibrium setting.

The first and most important is a standalone paper by Kaufmann et al. (2024), who develop a model essentially identical to that presented here but in a highly restricted environment.[3] Kaufmann et al. study the special case of an economy with only two goods, a linear supply curve, and symmetric consumers who have quasilinear utility in consumption and are indifferent to the supply of the quasilinear good.[4] In such a setting the quasilinear good is effectively money, so the model amounts to a partial equilibrium analog to the general equilibrium model explored here.

These restrictions allow the authors to illustrate some of the conceptual subtleties of strategic consumer behavior more straightforwardly, and to detail the economic and policy implications of ethical consumerism in the restricted setting. Their results however offer little guidance to a consumer hoping to achieve a given impact on production quantities in a world of interconnected markets. As outlined below, a contribution of the present paper is to demonstrate that many economic implications of ethical consumerism which can be identified in the partial equilibrium case do not generalize.

The second strand is the general literature on the private provision of public goods. Bergstrom et al. (1986) study equilibrium spending behavior by public good providers in light of crowd-out issues like those discussed above. The model we will consider is thus conceptually related to the model introduced by Bergstrom et al. and developed throughout the subsequent literature on the private provision of public goods and bads.

[3]That paper and this one were written roughly concurrently.

[4]In one setting they consider three goods, of which two are in a certain sense perfect substitutes.

Our model, however, will vary from the models typically explored by that literature in two interrelated ways. First, whereas models of public good provision games typically treat prices as exogenous, we will allow goods' prices to be determined endogenously by the starting endowments, production technologies, and quantities of other goods purchased. Second, we will allow some individuals to have preferences that depend only on their own consumption baskets and not at all on total supply levels. (Without the first variation—i.e. the introduction of endogenous prices—fully "selfish" individuals have no impact on the game played among those who do care about total supply levels. Selfish individuals can therefore safely be excluded from the model.)

The final strand is the literature, apparently confined so far entirely to agricultural economics, on "equilibrium displacement models", or EDMs. (See Wohlgenant (2011) for a review.) EDMs generate predictions about the impact that purchases of a particular good have on that good's equilibrium supply level. Norwood and Lusk (2011), for instance, estimate the price elasticities of supply and demand for various animal products and, from these estimations, calculate the extent to which marginal purchases of a given animal product change the equilibrium quantity supplied of that product. Wilkinson (2022) uses a similar EDM analysis in arguing that consumers are morally obligated to account for the market externalities of their purchases: i.e. the impacts that their purchases have on prices, and the impacts of these price-shifts on others' purchases and ultimately others' welfare.

The two pieces cited above—along with most of the EDM literature—only consider the impact of purchasing a certain good on the equilibrium supply of that good. In effect, they consider only the equilibrium supply-shift presented graphically in Section 1.1. For our purposes, such an analysis is relevant only in the event

that cross-price elasticities of demand and supply are zero across goods over whose supply a consumer has ethical preferences, and complete only under still stronger conditions. Goods about which one has ethical preferences, however, are often highly substitutable (for consumers and other producers) with goods about which one has similar ethical preferences. Recall the example of Section 1.1: a consumer who reduces her consumption of chicken on ethical grounds is likely to be concerned with decreasing the production not only of chicken but also of other meat products. Partial equilibrium analyses like the two cited above may thus be highly misleading.

An obscure but especially relevant step toward generality, within the EDM literature, is taken by Gardner (1987). Gardner's framework allows for an analysis of multiple goods simultaneously, in which changes to an individual's demand for one good affect the equilibrium prices and supply levels of other goods under consideration. His framework nonetheless falls short of a general equilibrium framework in two ways. First, it can serve only as a partial guide to the individual ethical consumer, as it breaks down when the consumer attempts to account for her impacts on all markets at once. Second, it serves *only* as a guide to the individual ethical consumer. It does not address the complexities in defining and finding the equilibria that arise when multiple (let alone all) consumers choose demands in light of their ethical preferences. It is thus in no way continuous with the literature on public good games, referenced above.

The general equilibrium analysis presented here fills this gap.

1.3 Outline

Section 2 introduces the model: a game in which consumers strategically choose demand functions to maximize preferences over consumption and supply, embedded

in a conventional general equilibrium setting. Section 3 characterizes the equilibria of the game, proposes a natural equilibrium refinement, and identifies conditions under which an equilibrium is guaranteed to exist.

Section 4 explores the often counterintuitive relationship between ethical preferences and optimal consumer behavior in the general equilibrium setting of Sections 2 and 3. The section thus highlights the importance to the ethical consumer of accounting for cross-market effects, and difficulty for small consumers—and, by extension, public good contributors—of predictably achieving desired effects on quantities produced.

Section 5 concludes with a discussion of the value of better tools for identifying such effects and a summary of how the analysis above might be extended.

2 Model

The model is one in which consumers choose demand functions in order to satisfy their preferences over both aggregate supply and their own consumption. We introduce the model in two stages. First, we define the economic environment within which the game, in which individuals choose their demand functions, will be played. This will allow us to define the economic outcomes that obtain given a profile of demand functions. Next we introduce preferences and define the game itself, along with the relevant equilibrium concepts.

2.1 Economic environment

Production and prices

There are $L \geq 2$ goods, indexed by ℓ or k.

There are I individuals, indexed by i, with endowment vectors $e^i > 0$.[5] The vector of total endowments is e.

A price vector p is of dimension L, with $p > 0$. Given prices p, the vector of net production $y(p)$ maximizes $p \cdot y$ within a production possibility set Y:

$$y(p) \equiv \underset{y \in Y}{\text{argmax}} \; p \cdot y. \tag{285}$$

Y is compact, and its efficient frontier is strictly concave, ensuring that $y(p)$ exists and is unique. Y satisfies $y \geq -e \; \forall y \in Y$, ensuring that supply $s(p) \equiv e + y(p)$ is non-negative. It follows from (285) that $y(\cdot)$ and thus $s(\cdot)$ are homogeneous of degree (h.o.d.) 0.

Definition 1. *An environment E is a tuple of endowments $\{e^i\}_{i=1}^I$, aggregate profit shares $\{\theta^i\}_{i=1}^I$, and a production function $y(\cdot)$ such that*

$$e^i > 0 \;\; \forall i; \tag{286}$$

$$\theta^i \geq 0 \;\; \forall i, \;\; \textstyle\sum_{i=1}^I \theta^i = 1; \; \textit{and} \tag{287}$$

$$y(p), \text{ defined on } \mathbb{R}^L_{++}, \text{ maximizes } p \cdot y \text{ subject to } y \in Y, \text{ where} \tag{288}$$

Y is compact, its efficient frontier is strictly concave, and $y \geq -e \; \forall y \in Y$.

(An economy is defined below as an environment and a profile of demand functions.)

We have assumed that production y maximizes $p \cdot y$, not that production maximizes *profits*. We will define profits below, finding that if consumers exhaust their budgets, aggregate profits equal $p \cdot y$. We will not in general assume that all con-

[5] When clear, we will let 0 denote a column vector of zeroes. When it is necessary to distinguish this vector from the scalar, we will denote the L-dimensional column vector of zeroes by $\mathbb{0}_L$. Given equal-length vectors b, c, we will use the notation that $b \geq c$ if $b_\ell \geq c_\ell \; \forall \ell$; $b > c$ if $b \geq c$ and $b \neq c$; $b \gg c$ if $b_\ell > c_\ell \; \forall \ell$; and $b \propto c$ if $b = \kappa c$ for some real κ.

sumers exhaust their budgets, however, and if some consumers do not exhaust their budgets, aggregate profits either exceed $p \cdot y$ or are undefined. Nevertheless, we will see that the production vector y that maximizes $p \cdot y$ also maximizes aggregate profits (when these are defined).

Demand

Aggregate profits (hereafter, profits) are denoted π. Profits are divided among consumers according to shares $\theta^i \geq 0$. Given prices p and profits π, i's budget equals $p \cdot e^i + \theta^i \pi$.

Individual i chooses a demand function, defined over positive prices and non-negative profits: $x^i(p, \pi)$. (Demand functions will be chosen strategically to satisfy preferences, as discussed in the next section. For now we will treat them as exogenous.) The aggregate demand function, or sum of individual demand functions, is denoted $x(\cdot)$.

Definition 2. *A function $x^i(p, \pi)$ from $\mathbb{R}^L_{++} \times \mathbb{R}_+$ to $\mathbb{R}^L_+$ is an admissible demand function for i in environment E if $x^i(\cdot)$ is h.o.d. 0, feasible in the sense that*

$$p \cdot x^i(p, \pi) \leq p \cdot e^i + \theta^i \pi \quad \forall (p, \pi) \in \mathbb{R}^L_{++} \times \mathbb{R}_+, \tag{289}$$

and non-decreasing in π.

We will denote the set of admissible demand functions for i in E by $\mathbb{A}^i(E)$ and the set of admissible demand function profiles in E by $\{\mathbb{A}^i(E)\}$. An admissible aggregate demand function is the sum of an admissible demand function profile, and the set of admissible aggregate demand functions is denoted $\mathbb{A}(E)$.

We deliberately impose only minimal restrictions on demand functions here in

order to emphasize the point that individuals may choose their demand functions strategically, and we have not yet determined what demand functions (if any) are incompatible with optimization in the current context. Further restrictions on demand functions will be introduced below where appropriate.

Profits

Given prices $p \gg 0$ and an admissible aggregate demand function $x(\cdot)$, profits equal

$$\pi(p, x(\cdot)) \equiv \underline{\pi \text{ such that } p \cdot x(p, \pi) = p \cdot (e + y(p))} \tag{290}$$

if the expression above is defined (and are undefined otherwise). By the feasibility of the individual demand functions,

$$p \cdot x(p, p \cdot y(p)) \leq p \cdot (e + y(p)). \tag{291}$$

If defined, therefore, π weakly exceeds $p \cdot y(p)$.

An admissible demand function profile does guarantee that profits are well-defined: $\pi(p, x(\cdot))$ may not exist or not be unique. Observe however that profits are guaranteed to be defined under the relatively weak condition that aggregate demand is nonsatiated, meaning that $p \cdot x(p, \pi)$ strictly increases without bound in π. Since $p \cdot x^i(p, \pi)$ weakly increases in π for all i, nonsatiation is guaranteed as long as $p \cdot x^i(p, \pi)$ strictly increases without bound in π for some i.

This definition of profits is an alternative to the more conventional $\pi(p, y) = p \cdot y$, and is commonly used in the event that some individuals may be satiated.[6]

[6]Konovalov (2005) offers a review of the literature using this approach, at least as of 2005. The resulting economic equilibria are called "dividend equilibria" or "Walrasian equilibria with slack". We will use the latter term, to emphasize that these equilibria generalize the familiar notion of Walrasian equilibrium.

Here, we do not assume local nonsatiation: we allow for the possibility that, in some circumstances and for some preferences individuals might have about aggregate supply levels, an individual maximizes his utility by choosing a demand function in which he does not always exhaust his budget. Defining profits to equal $p \cdot y$ would imply that, if any individual does not exhaust his budget, $p \cdot x < p \cdot s$. Firms would in effect remain in possession of unsold goods of value of $p \cdot (s - x)$, leaving them to rot on the shelves, even if some of their owners would have liked to consume them.

Definition (290), by contrast, amounts to the more natural and realistic assumption that any such excess production would belong to the producer, and thus to individuals according to shares θ. This implies that if $p \cdot x(p, \pi) < p \cdot s(p)$, the proposed profit level π is incompatible with $(p, x(\cdot))$. Given prices p and dividend payouts totaling π, profits inclusive of unsold goods in fact exceed π by $p \cdot (s(p) - x(p, \pi))$.

Observe that when individuals do exhaust their budgets,

$$p \cdot x(p, p \cdot y) = p \cdot e + p \cdot y = p \cdot s. \tag{292}$$

Profits then are always defined and equal $p \cdot y$, as usual. Also, even in the more general setting of (290), given that profits are uniquely defined, because $p \cdot x(p, \pi)$ weakly increases in π, $y(p)$ maximizes profits iff $y(p)$ maximizes $p \cdot s$ and thus $p \cdot y$.

Note that instead of introducing a set of profit-maximizing firms, we are positing that production maximizes profits subject to a single, abstract production technology. If we modeled firms explicitly, we would have to model how their behavior responded to the diverse preferences of their owners, not all of whom in this setting will care only about profit maximization. We would likewise have to explore the possibility that some distributions of firm ownership are incompatible with equilibrium. Individuals with preferences regarding the production levels of some goods might,

for instance, want to concentrate their capital holdings in a few relevant firms so as to wield influence as "activist investors". To restrict our focus to the equilibrium consequences of market purchases, we will therefore simply assume that production proceeds along profit-maximizing lines.

Economic equilibria

Definition 3. *Given an admissible and nonsatiated aggregate demand function* $x(p, \pi)$ *and a supply function* $s(p)$*, implicit demand equals*

$$\chi(p) \equiv x(p, \pi(p, x(\cdot))), \tag{293}$$

and excess demand equals

$$z(p) \equiv \chi(p) - s(p). \tag{294}$$

Let

$$\mathcal{I} \equiv [I_{L-1}, \mathbb{0}_{L-1}] : \tag{295}$$

the $(L-1) \times L$ matrix whose first $L-1$ columns constitute the $(L-1) \times (L-1)$ identity matrix and whose last column is the zero vector.

Definition 4. *Given price vector* $p \gg 0$*, the corresponding normalized price vector equals*

$$\hat{p} \equiv \mathcal{I}p/p_L. \tag{296}$$

The L^{th} entry of p/p_L is always 1, and left-multiplication by $\mathcal{I}$ removes this

redundant entry. In general, we will use the "hat" symbol to denote normalization from L to $L-1$ dimensions in this way, and we will use the upside-down hat (or "check") to denote expansion from $L-1$ to L dimensions, so that given normalized price vector $\hat{p}$,

$$\check{\hat{p}} \equiv (\hat{p}, 1) \ \big(= p/p_L, \text{ if } p \text{ is given}\big); \tag{297}$$

and normalized supply, excess demand, etc. functions are

$$\begin{aligned} \hat{s}(\hat{p}) &\equiv \mathcal{I}s(\check{\hat{p}}), \\ \hat{z}(\hat{p}) &\equiv \mathcal{I}z(\check{\hat{p}}), \text{ etc.} \end{aligned} \tag{298}$$

Definition 5. *Price vector $\bar{p} \gg 0$ is a Walrasian equilibrium with slack (WES) of environment E and aggregate demand function $x(\cdot) \in \mathbb{A}(E)$ [or demand function profile $\{x^i(\cdot)\} \in \{\mathbb{A}^i(E)\}$] if $z(\bar{p}) = 0$.*
Normalized price vector $\hat{\bar{p}}$ is a normalized WES (NWES) of $(E, x(\cdot))$ if $\hat{z}(\hat{\bar{p}}) = 0$.

Since in general profits may be undefined, so may implicit demand, and so may excess demand. If $\bar{p}$ is a WES of $(E, x(\cdot))$, therefore, it follows that $\pi(\bar{p}, x(\cdot))$ is defined.

Given an environment E, an aggregate demand function $x(\cdot) \in \mathbb{A}(E)$, and a WES $\bar{p}$ of $(E, x(\cdot))$, we will write as shorthand $\bar{\pi} \equiv \pi(\bar{p}, x(\cdot))$.

Definition 6. *A WES $\bar{p}$ of environment E and aggregate demand function $x(\cdot) \in \mathbb{A}(E)$ is a regular WES (RWES) if (a) the Jacobian $D\hat{z}(\hat{\bar{p}})$ is nonsingular and (b) $s(\cdot)$ is locally $\mathcal{C}^1$ around $\bar{p}$.*

We will refer to a normalized regular WES as an NRWES. The regularity of a Wal-

rasian equilibrium is typically defined to be condition (a) in isolation (see e.g. Mas-Colell et al. (1995), p. 591). We define regularity more strongly here because we will only ever require (a) in conjunction with (b).

Definition 7. *Given a WES $\bar{p}$ of environment E and demand function profile $\{x^i(\cdot)\}$, the demand function profile is locally regular if each $x^i(\cdot)$ is locally $\mathcal{C}^1$ and $\frac{\partial}{\partial \pi}\big(p \cdot x(p,\pi)\big) > 0$ around $(\bar{p},\bar{\pi})$.*

2.2 Strategic demand function choice

It may be helpful to open with a summary of the strategic interaction we will define precisely in this section.

An environment E obtains, and each individual i has a utility function $u^i(x^i, s)$, defined over his own consumption and aggregate supply. Each i chooses a demand function $x^i(\cdot) \in \mathbb{A}^i(E)$. In conjunction with the environment's supply function $s(\cdot)$, $\{x^i(\cdot)\} \in \{\mathbb{A}^i(E)\}$ is compatible with a (perhaps empty) set of RWES $\bar{p}$, which given $\{x^i(\cdot)\}$ uniquely determine a profit level, aggregate supply, and individual demands. A demand function profile $\{x^i(\cdot)\}$ and a compatible RWES $\bar{p}$ will be called an equilibrium if, for each i and each demand function $\tilde{x}^i(\cdot)$ admissible for i in E (subject to a smoothness condition),

- there is a unique normalized price vector $\hat{p}$ near $\hat{\bar{p}}$ that is an NRWES compatible with $\{x^{-i}(\cdot), \tilde{x}^i(\cdot)\}$ and $s(\cdot)$; and
- the u^i that obtains given $(\hat{p}, \{\tilde{x}^i(\cdot), x^{-i}(\cdot)\})$ is, in the limit as the economy is replicated and i's market power vanishes, no greater than the u^i that obtains given $(\hat{\bar{p}}, \{x^i(\cdot)\})$.

Preferences

Individual i has a utility function $u^i(\cdot)$ representing her preferences over $(x^i, s) \in \mathbb{R}^{2L}_+$. We will assume that her preferences are additively separable across x^i and s, so that it is meaningful to refer independently to consumption preferences ranking consumption baskets x^i and external preferences ranking supply baskets s, and that she has a utility function of the form

$$u^i(x^i, s) = v^i(x^i) + w^i(s). \tag{299}$$

Her consumption utility function $v^i(\cdot)$ is continuous. Her externality function[7] $w^i(\cdot)$ is $\mathcal{C}^1$.

Note that we do not here impose the standard assumptions that $v^i(\cdot)$ is locally nonsatiated (or even non-decreasing). In particular, we do not rule out the possibility that $v^i(\cdot)$ is independent of x^i_ℓ for some or all ℓ. Thus, though the application to ethical consumerism guides our running interpretation of the model, the framework allows for individuals to purchase some goods entirely because of the desired effects of such purchases on aggregate supply—i.e. as acts of philanthropy. It likewise allows for the interpretation that some agents are philanthropic foundations, exhibiting preferences defined exclusively over supply.

A utility function profile $\{u^i(\cdot)\}_{i=1}^I$ is admissible if $u^i(\cdot)$ satisfies the conditions above for all i.

Definition 8. *An economy $\mathcal{E}$ is an admissible utility function profile $\{u^i(\cdot)\}$ and an environment E.*

[7] These terms are intended only to distinguish the expression $v^i(\cdot)$ for i's preferences over the quantities of goods she herself purchases from the expression $w^i(\cdot)$ for i's preferences over aggregate supply levels. The latter will often consist primarily of concerns about the externalities that supply imposes on others, but as noted in Section 1, it need not do so.

Given an economy $\mathcal{E} = (\{u^i(\cdot)\}, E)$, an (individual/aggregate/etc.) demand function is "admissible in $\mathcal{E}$", denoted e.g. $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$, if the object in question is admissible in E. Likewise, $\bar{p}$ is a [N][R]WES of $(\mathcal{E}, x(\cdot))$ iff it is a [N][R]WES of $(E, x(\cdot))$.

WES displacement

Definition 9. *Given economy*

$$\mathcal{E} = \big(\{u^i(x^i, s)\}_{i=1}^I, \{e^i\}_{i=1}^I, \{\theta^i\}_{i=1}^I, y(p)\big),$$

i's n-replicated utility function equals

$$u^{i(n)}(x^i, s) \equiv v^i(x^i) + nw^i(s/n). \tag{300}$$

The n-replicated economy equals

$$\mathcal{E}^{(n)} \equiv \big(\{u^{[i \bmod I](n)}(x^i, s)\}_{i=1}^{nI}, \{e^{i \bmod I}\}_{i=1}^{nI}, \{\theta^{i \bmod I}/n\}_{i=1}^{nI}, ny(p)\big). \tag{301}$$

n-replicated (individual/aggregate/etc.) demand functions equal

$$\begin{aligned} x^{i(n)}(p, \pi) &\equiv x^i(p, \pi/n), \\ x^{(n)}(p, \pi) &\equiv nx(p, \pi/n), \\ x^{-i(n)}(p, \pi) &\equiv x^{(n)}(p, \pi) - x^{i(n)}(p, \pi), \text{ and} \\ \{x^i(p, \pi)\}^{(n)} &\equiv \{x^{i \bmod I}(p, \pi/n)\}_{i=1}^{nI}. \end{aligned} \tag{302}$$

n-replicated profits equal

$$\begin{aligned}\pi^{(n)}(p, x(\cdot)) &\equiv \pi : p \cdot x(p, \pi) = np \cdot s(p) \\ &= n\big(\pi : p \cdot x(p, \pi)/n = p \cdot s(p)\big) \\ &= n\pi(p, x(\cdot)/n). \end{aligned} \tag{303}$$

In (300), the utility function of i and her "replicas" in the n-replicated economy is defined so that replication diminishes i's market power without changing either the location of supply in her externality function or her preferences over marginal changes in supply.

If we adopt

$$u^{i(n)}(x^i, s) = u^i(x^i, s),$$

then replication does not change i's economic circumstances only by diminishing her market power, but also by changing the value of a variable over which her preferences are defined, namely supply. For instance if $w^i(s)$ is strictly decreasing and strictly concave in s_ℓ—as it might be if the production of good ℓ is emissions-intensive and i considers environmental damages to be convex in emissions—then the larger quantity of emissions in the replicated economy increases the intensity of i's preference for marginal emissions reductions.

If we adopt

$$u^{i(n)}(x^i, s) = u^i(x^i, s/n),$$

so that i in the n-replicated economy simply acts as if supply were $1/n$ times as large as it is, then the problem above is resolved, but a new one is introduced. An ethical consumer's concern in a given economy, large or small, is with the impact of a potential purchase on s itself—not its impact on s/n, an impact which is of course

only $1/n$ times as large. The impact of a given purchase on s is asymptotically constant and generically nonzero as an economy is replicated (as shown formally in the proof of Proposition 3 below, and illustrated by Figure 1), whereas the impact of a given purchase on s/n falls to zero.

Utility function (300) resolves both issues. Note that if each $w^i(\cdot)$ is CRS, our replication is defined in the usual way, without any modification to individuals' utility functions.

Proposition 1. ***Preservation of WES under replication***

Given economy $\mathcal{E}$ and aggregate demand function $x(\cdot) \in \mathbb{A}(\mathcal{E})$, $\bar{p}$ is a [N][R]WES of $(\mathcal{E}, x(\cdot))$ iff $\bar{p}$ is a [N][R]WES of $(\mathcal{E}^{(n)}, x^{(n)}(\cdot))$ $\forall n \geq 1$.

Proof. See Appendix B.1. □

Proposition 2. ***Local RWES under demand shifts***

Let $\bar{p}$ be an RWES of economy $\mathcal{E}$ and locally regular demand function profile $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}^i(\mathcal{E})\}$. Then there exists an $\epsilon > 0$ and an $\underline{n} \geq 1$ such that

a. for any i and $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$ that is $\mathcal{C}^1$ around $(\bar{p}, \bar{\pi})$, for $n \geq \underline{n}$ there exists an NRWES $\hat{p}$ of $(\mathcal{E}^{(n)}, \bar{x}^{-i(n)}(\cdot) + x^{i(n)}(\cdot))$ within the ϵ-neighborhood of $\hat{\bar{p}}$; and

b. within the ϵ-neighborhood of $\hat{\bar{p}}$, $\hat{p}$ is the unique NWES.

Proof. See Appendix B.2. □

Definition 10. *An economy $\mathcal{E}$ is <u>large</u> with respect to $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}^i(\mathcal{E})\}$ and WES $\bar{p}$ of $(\mathcal{E}, \bar{x}(\cdot))$ if Proposition 2 holds for $(\mathcal{E}, \bar{x}(\cdot), \bar{p})$ given $\underline{n} = 1$.*

Equilibrium concept

Given an economy $\mathcal{E}$, demand functions $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}^i(\mathcal{E})\}$, and an RWES $\bar{p}$ such that each $\bar{x}^i(\cdot)$ is locally $\mathcal{C}^1$ around $(\bar{p}, \bar{\pi})$, choose n such that $\mathcal{E}^{(n)}$ is large with respect

to $\{x^i(\cdot)\}$. Denote the locally unique NWES of $(\mathcal{E}^{(n)}, \bar{x}^{-i(n)}(\cdot) + x^{i(n)}(\cdot))$, for some locally $\mathcal{C}^1$ $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$, by

$$\hat{p}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big).$$

Then, given locally $\mathcal{C}^1$ demand function $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$, let $u_{\bar{p}}^{i(n)}\big(x^i(\cdot)\big)$ denote the utility i achieves by demand function $x^{i(n)}(\cdot)$ in the replicated economy and demand function profile:

$$u_{\bar{p}}^{i(n)}\big(x^i(\cdot)\big) \equiv u^i\Bigg(x^{i(n)}\Big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big), \pi^{(n)}\big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big), \bar{x}^{-i(n)}(\cdot) + x^{i(n)}(\cdot)\big)\Big), s\Big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big)\Big)\Bigg). \tag{304}$$

The expression appears complex because it captures the several channels through which a choice of demand function $x^i(\cdot)$ in economy n affects i's utility, but each channel is straightforward. In economy n, i's utility, given demand function $x^i(\cdot)$, depends in the usual way on his consumption x^i and supply s. The former depends on prices and profits, the latter only prices. Profits depend on prices and $x^i(\cdot)$. Prices depend on $x^i(\cdot)$.

We can now define a competitive equilibrium concept suited to the case in which consumers care about their impacts on supply.

Definition 11. *Let $\bar{p}$ be an RWES of economy $\mathcal{E}$ and locally regular demand function profile $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}(\mathcal{E})\}$. Then $(\bar{p}, \{\bar{x}^i(\cdot)\})$ is a competitive equilibrium with supply externalities (CESE) of $\mathcal{E}$ if, for all i and all locally $\mathcal{C}^1$ $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$,*

$$\lim_{n\to\infty}\Big[u_{\bar{p}}^{i(n)}(x^i(\cdot)) - u_{\bar{p}}^{i(n)}(\bar{x}^i(\cdot))\Big] \leq 0. \tag{305}$$

So $(\bar{p}, \{\bar{x}^i(\cdot)\})$ is a CESE if, in response to $\bar{x}^{-i}(\cdot)$, each i chooses an admissible and locally $\mathcal{C}^1$ $x^i(\cdot)$ that would be optimal for i if the economy were "infinitely large"

relative to i. In effect, $x^i(\cdot)$ must be a demand function i is content to choose if he acts as a price-taker (because his impacts on prices are infinitesimal) but not as a quantity-taker (because his impacts on prices affect the behavior of infinitely many other agents, resulting in positive quantity effects even in the limit).

After addressing certain technicalities, the strategic interaction described in this section can be expressed as a normal form game, and CESE may be understood as a weakening of Nash equilibrium. This is discussed in Appendix A.1.

2.3 Summary

The components of the model are summarized below. The number of goods $L \geq 2$ and individuals I is fixed throughout. Demand and utility functions are assumed to satisfy the "admissibility" conditions defined below unless otherwise stated.

Environment	(1a) Profiles of endowments $\{e^i\} > 0$ and (1b) profit shares $\theta^i \geq 0$ summing to 1; and (2) a production function $y(p)$ maximizing $p \cdot y$ within a production set Y (which satisfies certain conditions).

Quantities until the first horizontal line are defined given an environment.

[Admissible] demand function	A function $x^i(p, \pi)$ that is h.o.d. 0, feasible ($p \cdot x^i \leq p \cdot e^i + \theta^i \pi$), and non-decreasing in π.
Supply	Endowments plus production: $s(p) \equiv e + y(p)$.
Profits	Given prices p and agg. dmd. fcn. $x(p, \pi)$, π sets $p \cdot x(p, \pi) = p \cdot s(p)$. (Defined if $x(p, \pi)$ increases without bound in π.)

Implicit demand	Given $x(p, \pi)$, $\chi(p)$ expresses agg. dmd. as a function of prices, incorporating how prices determine profits.
WES	Prices p setting excess demand $\chi(p) - s(p)$ equal to 0.
RWES	A WES at which $s(\cdot)$ is $\mathcal{C}^1$ and the Jacobian of "normalized" excess dmd. (for $L-1$ goods, in their prices) is nonsingular.
Locally regular dmd. fcn. profile	Given WES p, $\{x^i(\cdot)\}$ such that each $x^i(\cdot)$ is locally $\mathcal{C}^1$ and $\frac{\partial}{\partial \pi}(p \cdot x(p, \pi)) > 0$ at $(p, \pi(p, x(\cdot)))$.
[Admissible] utility function	$u^i(x^i, s) = v^i(x^i) + w^i(s)$, with $v^i(\cdot)$ continuous and $w^i(\cdot)$ $\mathcal{C}^1$.
Economy	(1) An environment and (2) a utility function profile.
CESE	Given an economy, (1) an RWES and (2) a locally regular dmd. fcn. profile s.t. any utility gain to any i of deviating to another locally $\mathcal{C}^1$ dmd. fcn. vanishes as the economy is replicated.

3 Equilibrium

3.1 Characterization

Given an economy $\mathcal{E}$, define the following terms:

$$\begin{aligned}
\delta(p, x(\cdot)) &\equiv \text{The gradient of the aggregate Engel curve at } (p, \pi(p, x(\cdot))), \text{ i.e.} \\
&\quad \nabla_\pi x(p, \pi(p, x(\cdot))). \\
G(p, x(\cdot)) &\equiv \text{The generalized inverse } G \text{ of } -Dz(p), \text{ where } z(\cdot) \text{ is implied} \quad (306) \\
&\quad \text{by } (\mathcal{E}, x(\cdot)), \text{ with } G\delta(p, x(\cdot)) = \mathbb{0}_L \text{ and whose bottom row} \\
&\quad \text{equals } \mathbb{0}_L^T. \\
\psi^i(p, x(\cdot)) &\equiv \big(Ds(p)\, G(p, x(\cdot))\big)^T \nabla w^i(s(p)). \quad (307)
\end{aligned}$$

We can now characterize CESE relatively simply.[8]

Proposition 3. ***Characterization of CESE***

Let $\bar{p}$ be an RWES of economy $\mathcal{E}$ and locally regular demand function profile $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}(\mathcal{E})\}$. Then $(\bar{p}, \{\bar{x}^i(\cdot)\})$ is a CESE of $\mathcal{E}$ iff

$$\bar{x}^i(\bar{p}, \bar{\pi}) \in \underset{x^i}{\operatorname{argmax}} \left(v^i(x^i) + \psi^i(\bar{p}, \bar{x}(\cdot)) \cdot x^i \right) \;\Big|\; \bar{p} \cdot x^i \leq \bar{p} \cdot e^i + \theta^i \bar{\pi} \;\; \forall i. \quad (308)$$

Proof. See Appendix B.3. □

An intuition for the result is as follows.

Starting from a WES $\bar{p}$, suppose i adjusts his demand around $\bar{p}$ by a just-feasible dx^i: i.e., by a value dx^i satisfying $\bar{p} \cdot dx^i = 0$, or $dx^i \perp \bar{p}$. The impact of this

[8] Since economic price substitution data is often reported in terms of elasticities, an analog to (307) in elasticity terms is offered in Appendix A.2.

demand-adjustment on prices will, given the appropriate smoothness conditions and given that i is a negligible part of the economy, be linear in dx^i, and can therefore be represented by $dp = G dx^i$ for some matrix G. G must satisfy

$$-Dz\, G\, dx^i = dx^i \;\; \forall dx^i \perp \bar{p}. \tag{309}$$

That is, the equilibrium price impact dp of demand-shift dx^i must motivate shifts in supply ($Ds\, dp$) and in others' demands ($D\chi\, dp$) such that the gap between total supply and others' demands changes by precisely dx^i. Recall that $-Dz = Ds - D\chi$.

G must therefore be a generalized inverse of $-Dz$. Furthermore, the equilibrium price impact of a demand-shift along the gradient of the aggregate Engel curve must be zero: such a shift can be precisely accommodated by changing the profit rate, and thus others' demands, without affecting prices or supply. We thus have $G\delta = \mathbb{0}_L$. Finally, we make G uniquely defined by restricting the space of price vectors under consideration to those whose L^{th} entries equal 1 (or, more precisely, we restrict the space of marginal price-*changes* dp to those with $dp_L = 0$, so that p_L remains fixed at any given value), imposing that the bottom row of G equal zero. There is a unique generalized inverse of $Ds - D\chi$ satisfying these two conditions. The conditions thus identify G.

Then, i's marginal "ethical impact" of demand-shift dx^i, starting from a given WES, equals $\psi^i \cdot dx^i$, where

$$\psi^{i\,T} = \nabla w^i \cdot Ds\, G; \tag{310}$$

G converts the demand-shift to a price-shift, Ds converts the price-shift to a supply-shift, and ∇w^i converts the supply-shift to a marginal ethical impact from i's per-

spective. Transposing (310) yields (307).

Example

To illustrate the equilibrium concept, consider the economy $\mathcal{E}^{(n)}$, an arbitrary n-replication of the following economy $\mathcal{E}$ with $I = 2$ individuals and $L = 2$ goods:

$$\begin{aligned} e^i &= (3/2, 0)^T \ \ \forall i, \\ \theta^i &= 1/2 \ \ \forall i, \\ v^1(x^1) &= \ln(x_1^1) + x_1^1/3 + x_2^1, \\ v^2(x^2) &= \ln(x_1^2) + x_2^2; \\ w^1(s) &= s_1 + 2s_2, \\ w^2(s) &= 0; \end{aligned} \tag{311}$$

and a production function

$$y(p) = \left(1 - 2\sqrt{\frac{p_2}{p_1}},\ 4 - 2\sqrt{\frac{p_1}{p_2}}\right) \tag{312}$$

as long as $p_1/p_2 \in [1/4, 4]$ to ensure an interior solution. Supply is therefore

$$s(p) = \left(4 - 2\sqrt{\frac{p_2}{p_1}},\ 4 - 2\sqrt{\frac{p_1}{p_2}}\right), \tag{313}$$

as long as $p_1/p_2 \in [1/4, 4]$.

Let $\bar{p} \equiv (1,1)$, $\bar{\psi}^1 \equiv (-1/3, 0)$, and $\bar{\psi}^2 \equiv (0,0)$, and let

$$\bar{x}^i(p,\pi) \equiv \underset{x^i}{\text{argmax}}\ \tilde{u}^i(x^i) \ \Big|\ p \cdot x^i \leq p \cdot e^i + \theta^i \pi \ \ \forall i \tag{314}$$

$$\text{where} \quad \tilde{u}^i(x^i) \equiv v^i(x^i) + \bar{\psi}^i \cdot x^i = \ln(x_1^i) + x_2^i \ \ \forall i \tag{315}$$

$$\Longrightarrow\ \bar{x}_1^i(p,\pi) = \min\left(\frac{3p_1 + \pi}{2p_1}, \frac{p_2}{p_1}\right) \ \ \forall i, \tag{316}$$

$$\bar{x}_2^i(p,\pi) = \max\left(0, \frac{3p_1 + \pi}{2p_2} - 1\right) \ \ \forall i. \tag{317}$$

Note that the specified externality weight of $\bar{\psi}_1^1 = -1/3$ cancels out the $+x_1^1/3$ term of (311), leaving the maximand $\tilde{u}^i(x^i)$ equal for both i. The individuals' demand functions are then identical because the individuals also have the same endowments and profit shares.

We will now show that $(\bar{p}, \{\bar{x}^i(\cdot)\}^{(n)})$ is a CESE of $\mathcal{E}^{(n)}$.

To take care of the preliminaries, first observe that demand function profile $\{\bar{x}^i(\cdot)\}$ is admissible in $\mathcal{E}$. As noted following Definition 9, it follows that $\{\bar{x}^i(\cdot)\}^{(n)}$ is admissible in $\mathcal{E}^{(n)}$. Observe also that $\{\bar{x}^i(\cdot)\}$, and thus $\{\bar{x}^i(\cdot)\}^{(n)}$, is locally $\mathcal{C}^1$ around $(\bar{p}, \pi)$ for any $\pi \geq 0$.

Next, let us confirm that $\bar{p}$ is a WES of $\mathcal{E}$ and demand function profile $\{\bar{x}^i(\cdot)\}$. From (312), $y(\bar{p}) = (-1, 2)$. Also, $\tilde{u}^i(x^i)$ is strictly increasing for both i, so both individuals exhaust their budgets at any p, π. Thus $\bar{\pi} \equiv \pi(\bar{p}, \bar{x}(\cdot)) = \bar{p} \cdot y(\bar{p}) = 1$. $\bar{x}(\bar{p}, \bar{\pi}) = s(\bar{p}) = (2,2)$, confirming that $\bar{p}$ is a WES.

Then, by substituting $p \cdot y(p)$ for π in (317) and summing across i, we have aggregate implicit demand $\bar{\chi}(p)$. From here, though we will not work through the details, it is straightforward to find excess demand $\bar{z}(\cdot)$ $(\equiv \bar{\chi}(\cdot) - s(\cdot))$. With normalized excess demand $\hat{\bar{z}}(\cdot)$ defined as in Definition 4, we can then confirm that $\bar{p}$ is regular

by Definition 6.

Finally, by Proposition 1, it follows that $\bar{p}$ is also an RWES of $(\mathcal{E}^{(n)}, \{\bar{x}^i(\cdot)\}^{(n)})$.

What makes the proposed $(\bar{p}, \{\bar{x}^i(\cdot)\})$ a CESE, beyond the fact that $\bar{p}$ is an RWES of $(\mathcal{E}, \{x^i(\cdot)\})$, is that $\bar{\psi} = \psi(\bar{p}, \bar{x}(\cdot))$. This can be calculated from the Definition (307) of $\psi(\cdot)$ (and the prior Definition 306 of $G(\cdot)$), but the identity may be understood as follows.

Let $\tilde{G}^{(n)}(p, x(\cdot))$ denote an $L \times L$ (here, 2×2) matrix mapping individual demand-changes into changes to equilibrium prices, in economy $\mathcal{E}^{(n)}$, given initial conditions in which aggregate demand is $x(\cdot)$ and the price vector p is an RWES of $(\mathcal{E}^{(n)}, x^{(n)}(\cdot))$.[9] Since excess demand is h.o.d. 0, we can without loss of generality fix p_L, i.e. require that $\tilde{G}(\cdot)$ have a bottom row of zeroes.

Now observe that $\bar{x}_2^i(\bar{p}, \bar{\pi}) > 0 \ \forall i$. The quasilinearity of (315) guarantees that, at $(\bar{p}, \bar{\pi})$, each i's marginal purchases are exclusively of good 2. Therefore

$$\delta(\bar{p}, \bar{x}(\cdot)) = (0, 1/\bar{p}_2)^T = (0, 1)^T. \tag{318}$$

This in turn implies that if one i reduces x_2^i, this simply increases profits, which others spend entirely on good 2. Supply levels do not change, and neither do prices. (A price-change would induce a supply-change.) In other words, $G(\bar{p}, \bar{x}(\cdot))(0, -1)^T = (0, 0)^T$. The upper-right entry of $G(\bar{p}, \bar{x}(\cdot))$ therefore equals zero.

The only remaining unknown entry of $G(\bar{p}, \bar{x}(\cdot))$ is its upper-left, representing the extent to which marginal purchases of good 1 increase the price of good 1. To find

[9]The tilde distinguishes $\tilde{G}^{(n)}(\cdot)$ from what would more naturally be denoted $G^{(n)}(\cdot)$—i.e. the analog of $G(\cdot)$ in economy $\mathcal{E}^{(n)}$—but which is independent of n. The latter maps marginal individual demand-changes dx^i to what marginal equilibrium price-changes *would* be if i also responded to the marginal price-change she herself induced. This distinction vanishes as n grows.

it, we will consider a marginal individual demand-change proportional to $(1, -1)^T$. Since $\bar{p}_1 = \bar{p}_2$, this demand-change by i is orthogonal to $\bar{p}$.

By $s(p)$ from (313), $\partial s_1(p)/\partial p_1 = 1$ and $\partial s_2(p)/\partial p_1 = -1$. That is, recalling that supply in $\mathcal{E}^{(n)}$ equals $ns(p)$, each marginal unit increase in p_1 from the $\bar{p}$ baseline induces an n-unit increase in the supply of good 1 and an n-unit decrease in the supply of good 2 in economy $\mathcal{E}^{(n)}$.

Likewise, by (316)–(317), each marginal unit increase in p_1 induces a 1-unit decrease in demand for good 1 by all $2n - 1$ consumers other than i. (Holding profits fixed, it also induces a 1-unit increase in demand for good 2. We have not shown that profits will in fact remain fixed as prices change; but we know that profit-changes will not affect demand for good 1, as long as all consumers are consuming a positive quantity of good 2.)

To maintain market clearing, p_1 must rise by just enough to induce an increase in s_1, and a decrease in x_1^{-i}, which sum to 1: the additional unit of good 1 which consumer i has resolved to buy. That is, we must have

$$\begin{aligned} n\,dp_1 + (2n-1)(dp_1) &= 1 \\ \implies dp_1 &= \frac{1}{3n-1}. \end{aligned}$$

As noted above, changes in demand for good 2, starting from the baseline of $(\bar{p}, \bar{x}(\cdot))$, have no impact on supply levels. It follows that $\psi_2^1(\bar{p}, \bar{x}(\cdot)) = \psi_2^2(\bar{p}, \bar{x}(\cdot)) = 0$, as desired. Moreover, since $w^2(s) = 0$, $\psi_1^2(\bar{p}, \bar{x}(\cdot)) = 0$ by (307), as desired. All that remains is to show that $\psi_1^1(\bar{p}, \bar{x}(\cdot)) = -1/3$.

Recall that the marginal individual demand-shift of $(1, -1)$ induces a marginal price-shift of $dp_1 = \frac{1}{3n-1}$. This price-shift, in turn, induces a marginal $\frac{n}{3n-1}$-unit

increase in the equilibrium supply of good 1 and $\frac{n}{3n-1}$-unit decrease in the equilibrium supply of good 2. Recalling that $w^1(s) = s_1 + 2s_2$, the marginal ethical impact of marginal individual demand-shift $(1, -1)$, from the perspective of consumer 1 (or any of her clones), equals $-\frac{n}{3n-1}$. As $n \to \infty$, this ethical impact approaches $-1/3$. By definition, therefore, $\psi_1^1(\bar{p}, \bar{x}(\cdot)) - \psi_2^1(\bar{p}, \bar{x}(\cdot)) = -1/3$. But $\psi_2^1(\bar{p}, \bar{x}(\cdot)) = 0$. Thus $\psi_1^1(\bar{p}, \bar{x}(\cdot)) = -1/3$, as desired.

With $\bar{\psi} = \psi(\bar{p}, \bar{x}(\cdot))$, $(\bar{p}, \bar{x}(\cdot))$ is a CESE of $\mathcal{E}$ and any of its replications.[10]

As this example illustrates, the impacts of consumer behavior after accounting for general equilibrium effects can differ substantially from the impacts one finds when one entirely ignores substitution by other parties. Inspecting $w^1(s)$ alone, one might expect that consumer 1 assigns an externality weight of 1 to purchasing a unit of good 1, and an externality weight of 2 to purchasing a unit of good 2, on any margin. Here, by contrast, we find that she assigns a negative weight to good 1 and a weight of 0 to good 2.

Likewise, external impacts in general equilibrium can differ substantially from those found after accounting only for partial equilibrium effects. Expressions (313) and (316) record an upward-sloping supply curve and a downward-sloping demand curve for both goods, respectively, around $p = (1, 1)$. A partial equilibrium analysis, like that of Kaufmann et al. (2024), would therefore conclude that, from individual 1's perspective, the externality weight of purchasing a unit of good 1 given prices $(1, 1)$ lay in $(0, 1)$, and that of purchasing a unit of good 2 lay in $(0, 2)$.

[10] $(\bar{p}, \bar{\psi})$ is also an RCESE of $\mathcal{E}$ and its replications, where RCESE is a refinement of CESE defined in Section 3.2 below. Furthermore, though we have not shown this here, this RCESE is unique.

Discussion

Proposition 3 states that each consumer i is indifferent between all demand functions which demand, at the equilibrium price and profit level, the basket x^i that maximizes $v^i(x^i) + \psi^i(\bar{p}, \bar{x}(\cdot)) \cdot x^i$ subject to her budget constraint. The proposition thus offers little guidance as to what sorts of behavior we might expect to see in equilibrium. The (cross-)price elasticities of demand i chooses *around* the equilibrium price and profit level are of no consequence for i, but because they affect $x(\cdot)$ and thus ψ^j for $j \neq i$, they affect other consumers' best-response demand functions. That is, in a CESE, consumers' demands are highly sensitive to their fellow consumers' arbitrary choices of threatened out-of-equilibrium behavior. Exotic behavior may therefore be motivated in equilibrium by mutually best-responding threats with no basis in anyone's preferences over supply or own consumption. A refinement of CESE, designed to rule out these "non-credible threats", will be discussed in Section 3.2.

Even in the absence of such a refinement, however, Proposition 3 offers complete guidance as to what *baskets* an individual in a large economy, with given preferences and a given endowment, should be content to buy at given prices and profits. Again, the individual should buy a basket x^i that maximizes $v^i(x^i)+\psi^i(\bar{p}, \bar{x}(\cdot))\cdot x^i$ subject to her budget constraint. If our goal is only to advise informed consumers with given ethical preferences, therefore, our analysis of equilibrium can end here. Further analysis of the impacts of ethical concerns on an ethical consumer's behavior may be found in Section 4.

3.2 Refinement

Let us refer to ψ^i as i's externality vector. Let ψ denote the $L \times I$ externality matrix whose column i equals ψ^i.

Given ψ^i, let us refer to

$$\tilde{u}^i_{[\psi^i]}(x^i) \equiv v^i(x^i) + \psi^i \cdot x^i$$

as i's quasi-utility function. $\tilde{u}^i_{[\psi^i]}(\cdot)$, for any $\psi^i \in \mathbb{R}^L$, represents preferences over all-things-considered preferences over purchasing choices $x^i \in \mathbb{R}^L_+$. These preferences incorporate both the private benefits of consuming a given basket and the external impacts of the production-changes induced by purchasing that basket.

As discussed at the end of Section 3.1, we would like a refinement of CESE in which individuals choose demand functions that behave reasonably not only at the precise $(\bar{p}, \bar{\pi})$ that obtain in economic equilibrium but at all (p, π), or at least all (p, π) near $(\bar{p}, \bar{\pi})$. Such demand functions can be motivated by giving each individual uncertainty about the prices and profits she will face.

Unfortunately, how an individual best responds to a shock to prices and profits depends on the source of the shock. This is because the external impacts of a given consumption decision may depend on the economic conditions determining prices and profits.

External impacts are approximately independent of the shocks to the prices and profits an individual faces if these shocks do not significantly affect supply or others' budget sets across the economy at large. They are approximately independent, for instance, if he faces idiosyncratic costs on the way to the shop, or his local shop faces idiosyncratic costs in stocking a good on a particular day. He then does best by choosing x^i to maximize $\tilde{u}^i_{[\psi^i]}$ subject to a budget constraint that varies with p and π.

If the shock to (p, π) stems from a shock to production, on the other hand,

our individual can anticipate that the supply vector now lies at a location in his externality function $w^i(\cdot)$ different from the location that would have been realized absent the shock, and this may alter ∇w^i. The shock to supply may also shift the value of Ds. Finally, others' demands will adjust for the same reasons as his own, potentially shifting $D\chi$. For all three reasons, in the event of a shock, the individual does best to demand the basket that maximizes some

$$v^i(x^i) + \psi^i(p, \pi) \cdot x^i,$$

where not only the budget constraint but also the externality vector ψ^i varies with p and π.

Finally, if the shock to (p, π) stems from a shock to the tastes of individuals other than i, the appropriate adjustment to ψ^i depends on the details of the shock. Without knowing the distribution of tastes and individuals from which the shocks are drawn—and certainly without knowing whether the observed (p, π) shocks stem from taste or supply shocks—observing dp and $d\pi$ is not enough to determine the appropriate adjustment, if any, to ψ^i. A complete treatment of out-of-equilibrium demands would require individuals to choose demand functions that depend not only on prices and profits but also, directly, on a third variable indexing the realization of the shock to production and/or tastes.[11]

This reveals the difficulty of defining reasonable behavior by ethical consumers in the event that prices or profits are out of equilibrium (absent strong linearity assump-

[11]Kaufmann et al. (2024) avoid these difficulties by imposing that individuals choose demand functions that not only are locally $\mathcal{C}^1$—i.e. have locally continuous derivatives—but have locally *constant* derivatives; and by assuming a supply function and symmetric externality functions such that, in our notation, Ds globally constant and Dw^i is globally constant for all i. They then observe a sense in which, under these conditions, optimal demand-shifts given supply shocks are approximately linear in the induced price-shifts, as desired.

tions, as detailed in footnote 11). For simplicity, and perhaps (not unrelatedly) for realism, we will choose the refinement in which individuals understand externality weights to be independent of price and profit shocks.

Definition 12. *A robust CESE (RCESE) is a CESE* $(\bar{p}, \{x^i(\cdot)\})$ *in which each* $x^i(\cdot)$ *maximizes* $\tilde{u}^i_{[\psi^i]}$ *at all* $p \gg 0, \pi \geq 0$.

Note that, if $w^i(\cdot) = 0$, $\psi^i(\cdot) = 0$ as well. One desirable feature of RCESE is therefore that individuals without ethical preferences adopt the demand functions that globally maximize $v^i(\cdot)$, as individuals are assumed to do in the conventional Walrasian setting. CESE, by contrast, requires only that an individual with $w^i(\cdot) = 0$ adopt a demand function that maximizes $v^i(\cdot)$ at equilibrium prices and profits.

RCESE is therefore a relatively straightforward generalization of Walrasian equilibrium, after technicalities around regularity, to the case in which individuals have preferences over total supply levels. This will also be stated formally in Proposition 4, in the section below.

After this subsection, we will restrict our attention to RCESE. Under some conditions, this lets us reframe the model in simpler terms.

Instead of positing that each i chooses an arbitrary admissible demand function $x^i(\cdot)$, we can posit that i chooses a vector ψ^i. If, given ψ, a unique demand function for each i maximizes $\tilde{u}^i_{[\psi^i]}(x^i)$, we may define

$$x^i_{[\psi^i]}(p, \pi) \equiv \underset{x^i}{\operatorname{argmax}} \ \tilde{u}^i_{[\psi^i]}(x^i) \ \Big| \ p \cdot x^i \leq p \cdot e^i + \theta^i \pi,$$ [12]

[12] Observe that $x^i_{[\psi^i]}(\cdot)$ is admissible wherever it is defined.

and let $x_{[\psi]}(\cdot)$ denote the sum of individual $x^i_{[\psi^i]}(\cdot)$. With

$$\pi(p, \psi) \equiv \pi(p, x_{[\psi]}(\cdot)), \tag{319}$$

we can then define the implicit individual demand function compatible with a given ψ as

$$\chi^i_{[\psi]}(p) \equiv \underset{x^i}{\operatorname{argmax}} \ \tilde{u}^i_{[\psi^i]}(x^i) \ \Big| \ p \cdot x^i \leq p \cdot e^i + \theta^i \pi(p, \psi),$$

and let

$$\chi_{[\psi]}(p) \equiv \sum_{i=1}^{I} \chi^i_{[\psi]}(p),$$
$$z_{[\psi]}(p) \equiv \chi_{[\psi]}(p) - s(p).$$

As with $\pi(\cdot)$ in (319), we can define

$$\delta(p, \psi) \equiv \delta(p, x_{[\psi]}(\cdot)),$$

and analogously extend (307) and the preceding terms—defined there as functions of admissible aggregate demand functions $x(\cdot)$—to be defined with externality matrices, and not only explicit aggregate demand functions, in their second arguments. Finally, we can define (p, ψ) to be an RCESE if $(p, \{x^i_{[\psi^i]}(\cdot)\})$ is an RCESE:

Definition 13. *Given $\bar{\psi}$, let $\bar{p}$ be an RWES of an economy $\mathcal{E}$ and a demand function profile $\{x^i_{[\bar{\psi}^i]}(\cdot)\}$ that is [defined and] locally regular around $(\bar{p}, \bar{\pi})$. Then $(\bar{p}, \bar{\psi})$ is an*

RCESE of $\mathcal{E}$ if

$$\psi(\bar{p}, \bar{\psi}) = \bar{\psi}.$$

3.3 Existence

As noted in Section 3.2, if there are no "ethical consumers", then RCESE is essentially equivalent to Walrasian equilibrium, subject to certain technicalities. We can state this result more formally:

Proposition 4. ***RCESE generalizes regular Walrasian equilibrium***
Let $\bar{p}$ be a Walrasian equilibrium of an economy $\mathcal{E}$ with $w^i(\cdot) = 0 \; \forall i$. If $\bar{p}$ is regular and $\{x^i_{[0]}(\cdot)\}$ is locally regular around $(\bar{p}, \bar{p} \cdot y(\bar{p}))$, then $(\bar{p}, \mathbb{0}_{L \times I})$ is an RCESE of $\mathcal{E}$.

Proof. Because $\bar{p}$ is a regular Walrasian equilibrium, $\bar{p}$ is an RWES (with $\bar{\pi} = \bar{p} \cdot y(\bar{p})$). Because each $w^i(\cdot) = 0$, each $\psi^i(\cdot) = \mathbb{0}_L$, by (307). Thus $\psi(\bar{p}, \mathbb{0}_{L \times I}) = \mathbb{0}_{L \times I}$. □

Determining when an RCESE exists in a less trivial setting is difficult, but existence can be proven under certain conditions.

Let Δ^{L-1} denote the simplex in $\mathbb{R}^L$. Given a price vector p, let

$$\tilde{p} \equiv p / (\mathbb{1}_L \cdot p) \in \Delta^{L-1}$$

denote its simplex normalization.

For a set A, let A° denote its interior and A^c denote its boundary.

Definition 14. *A supply function $s(\cdot)$ is uniformly smooth on a compact, convex set of price vectors $P \subset \Delta^{(L-1)\circ}$ if*

1. $s(p) \gg 0$ *iff* $\tilde{p} \in P^\circ$ *and*

2. $s(\cdot)$ *is* $\mathcal{C}^1$ *on* P *with* $D\hat{s}(\hat{p})$ *nonsingular for all* $\hat{p}$ *with* $\hat{\tilde{p}} \in P^\circ$,

where the interior and boundary of P *are defined with respect to the space* Δ^{L-1}.

For intuition, in two goods, the supply possibility set corresponding to a uniformly smooth supply function must resemble Figure 2.

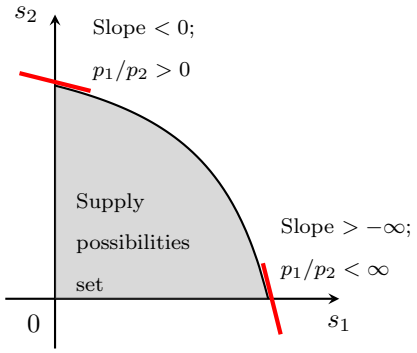


Fig. 2: Uniformly smooth supply in two goods

The relative price of good 1 p_1/p_2 motivating supply at a point s on the frontier is the negative of the slope of the tangent of the frontier at s (which is defined, by (iii)). The strict concavity of the frontier ensures that profit-maximizing production is unique at all prices. The compact range of slopes from the upper left to the bottom right ensures that, within this range of prices, supply is interior, and at or beyond its ends, the supply of some good is zero.

For $\mathcal{L} \subset \{1, ..., L\}$, let $\Delta^{L-1}_{\mathcal{L}}$ denote the face of Δ^{L-1} with entries of zero outside $\mathcal{L}$.

Definition 15. *A supply function* $s(\cdot)$ *is* <u>*quasi-monotonic*</u> *on a set of price vectors* P *if, for any* $\mathcal{L} \subset \{1, ..., L\}$ *and* $d \in \Delta^{L-1}_{\mathcal{L}}$,

$$p \in \operatorname*{argmax}_{p' \in P} p' \cdot d \implies s_\ell(p) > 0 \text{ for some } \ell \in \mathcal{L}.$$

That is, quasi-monotonicity on P ensures that if within P prices p maximize a weighted average of the prices of some set of goods $\mathcal{L}$, then the supply of at least one good in $\mathcal{L}$ is positive.

Proposition 5. ***Existence of RCESE given uniform smoothness and quasi-monotonicity of supply, separability of consumption utility, and symmetric consumers***

Suppose that $s(\cdot)$ is uniformly smooth and quasi-monotonic on a compact, convex set of price vectors $P \subset \Delta^{(L-1)\circ}$. Suppose also that individuals have symmetric utility functions $u^i(\cdot) = u(\cdot)$, endowments $e^i = e/I$, and profit shares $\theta^i = 1/I$; and that consumption utility is additively separable,

$$v(x^i) = \sum_{\ell=1}^{L} v_\ell(x_\ell^i), \tag{320}$$

with $v_\ell(\cdot)$ $\mathcal{C}^2$, strictly increasing, and satisfying

$$v_\ell''(x_\ell^i) < 0, \quad \lim_{x_\ell^i \to 0} v_\ell'(x_\ell^i) = \infty \ \ \forall \ell.$$

Then an RCESE $(\bar{p}, \bar{\psi})$ exists, with $\tilde{\bar{p}} \in P^\circ$, and $\bar{\psi}^i$ and x^i the same for all i.

Proof. See Appendix B.4. □

Proposition 6. ***Existence of RCESE given uniform smoothness of supply and separability and quasilinearity of consumption utility***

Suppose that:

1. *$s(\cdot)$ is uniformly smooth on a compact, convex set of price vectors $P \subset \Delta^{(L-1)\circ}$.*
2. *Either*

(a) $s(\cdot)$ is quasi-monotonic on P or

(b) for all ℓ, $\exists i : w^i(\cdot) = 0, e^i_\ell > 0$.

3. *For all i, $v^i(\cdot)$ is additively separable and quasilinear in good L,*

$$v^i(x^i) = \sum_{\ell=1}^{L-1} v^i_\ell(x^i_\ell) + v^i_L x^i_L, \tag{321}$$

with $v^i_\ell(\cdot)$ $\mathcal{C}^2$, strictly increasing, and satisfying $v''_\ell(x^i_\ell) < 0$ and $\lim_{x^i_\ell \to 0} v'_\ell(x^i_\ell) = \infty$ for all $\ell < L$.

Then for each i there is a $\underline{B}^i_L > 0$ and a $\underline{v}^i_L \geq 0$ such that if, for all i,

1. *$v^i_L > \underline{v}^i_L$ and*

2. *$p \cdot e^i_L + \theta^i p \cdot y(p) \geq \underline{B}^i_L \; \forall p \in P$,*

then an RCESE $(\bar{p}, \bar{\psi})$ exists, with $\bar{\psi}^i_L = 0 \; \forall i$, $\bar{\psi}^i_\ell < v^i_L \; \forall i$, and $\tilde{\bar{p}} \in P^\circ$.

Proof. See Appendix B.5. □

Existence more generally

The conditions required for Propositions 4–6 are highly restrictive because it is difficult to create economic conditions that guarantee the existence of an RWES throughout a large space of potential values of ψ. Well-known results establish that economies "generically" have regular and indeed only regular Walrasian equilibria, in the sense that singular Jacobians of excess demand around equilibrium price vectors can be made nonsingular with arbitrary perturbations to the economic primitives (see e.g. Mas-Colell et al. (1995), Proposition 17.D.5). In the current setting, however, genericity results of this kind are not obviously useful, because ψ is endogenous

to prices and economic primitives, and may in principle adjust to any perturbations so as to push equilibria toward regularity.

Despite this obstacle to identifying flexible conditions guaranteeing the existence of RCESE, it can easily be shown that RCESE exist far beyond the narrow conditions above. Given an economy $\mathcal{E} = (E, \{u^i(x^i)\})$, let $\bar{p}$ be an RWES of $(E, \{x^i_{[0]}(\cdot)\})$. Let $\psi = \psi(\bar{p}, \mathbb{0}_{L\times I})$, so that ψ is the best-response externality matrix at $\bar{p}$ to $\psi = \mathbb{0}_{L\times I}$. Finally, let

$$\mathcal{E}(\psi) \equiv (E, \{\tilde{v}^i_{[\psi^i]}(\cdot) + w^i(\cdot)\}),$$

where

$$\tilde{v}^i_{[\psi^i]}(x^i) \equiv v^i(x^i) - \psi^i \cdot x^i. \tag{322}$$

That is, let $\mathcal{E}(\psi)$ denote the economy with the same environment as $\mathcal{E}$, and whose individuals have the same externality functions $\{w^i(\cdot)\}$ as in $\mathcal{E}$, but who derive ψ^i_ℓ fewer units of consumption utility from good ℓ at all margins. Then, defining $\{x^i_{[\cdot]}(\cdot)\}$ with respect to the modified consumption utility functions, so that

$$x^i_{[0]}(p, \pi) \equiv \underset{x^i}{\operatorname{argmax}}\ \tilde{v}^i_{[\psi^i]}(x^i) \ \Big|\ p \cdot x^i \leq p \cdot e^i + \theta^i y(p),$$

$(\bar{p}, \psi)$ is an RCESE of $(\mathcal{E}(\psi), \{x^i_{[0]}(\cdot)\})$. This is because if each i's consumption utility is given by $v^i(x^i) - \psi^i \cdot x^i$ and she adopts externality weights ψ^i, then her quasi-utility function is simply $v^i(\cdot)$: just as it is if her consumption utility is $v^i(\cdot)$ and she adopts externality weights 0. The best response externality weights in the first case thus equal ψ, just as in the second.[13,14]

[13]The example of Section 3.1 can be constructed in this way, beginning from the economy in which $v^i(\cdot) = \ln(x^i_1) + x^i_2\ \forall i$.

[14]If $\psi^i_\ell < 0$ for some i, ℓ, $\tilde{v}_{[\psi^i]}(\cdot)$ as defined by (322) may decrease in x^i_ℓ on some margins

Multiplicity

Under standard assumptions, including ours, if consumers are symmetric or have quasilinear utility functions then there is a unique normalized Walrasian equilibrium. One might therefore wonder whether uniqueness extends to RCESE in the context of Propositions 5 and 6.

In general, it does not. Kaufmann et al. (2024) study a narrower setting in which consumers are symmetric *and* have well-behaved quasilinear consumption utility functions.[15] Indeed they also assume that there are only two goods (let us stipulate that the second is the quasilinear good); that the supply of the first increases linearly in its relative price; and that, after translating notation, the common externality function $w(\cdot)$ takes the form $w(s) = -\kappa s_1$, so that it is independent of the quasilinear good and linearly decreases in the other. These restrictions ensure that the relationship between the externality function and the externality weights is well-behaved: in particular, in any RCESE, $\psi_2^i = 0$ and $\psi_1^i \equiv \psi_1 \in (-\kappa, 0]$ is the same for all i. Even so, however, Kaufmann et al. find that there may be multiple RCESE.

For intuition, higher absolute values of ψ_1 motivate individuals to consume less of good 1. If they strongly prefer to consume at least a small quantity of good 1, their demand may be less responsive to price-changes once their consumption of good 1 has fallen. Lower price-responsiveness of demand (in combination with the constant price-responsiveness of supply) then implies that if a given individual chooses to consume more of good 1, the impact on the supply of good 1 is greater. Both low

even if $v^i(\cdot)$ is increasing. As shown in the proof of Proposition 9 (Appendix B.6), a profile of consumption utility functions compatible with RCESE $(\bar{p}, \bar{\psi})$ that preserves monotonicity can also be constructed.

[15] Strictly speaking, because they allow the supply and consumption of goods to be negative, their setting is not narrower than those considered here. Their multiplicity result however does not rely on this generalization.

and high absolute values of ψ_1 may therefore be self-confirming.

Kaufmann et al. also study the welfare implications of this multiplicity in their setting. In particular they note that equilibria in which $|\psi_1|$ is higher, and thus the consumption of good 1 is lower, Pareto-dominate those in which $|\psi_1|$ is lower. An implication is that the first fundamental theorem of welfare economics does not generally hold when individuals have preferences over supply, even if these preferences are identical and motivate consumers to shift their demands to some extent in the collectively desired direction. This is unsurprising in light of the close analogy between the game of strategic demand choice, in which individuals have preferences over supply, and a traditional public good game.

We will not explore uniqueness further here. The null uniqueness results cited above suffice to show that assumptions guaranteeing the uniqueness of Walrasian equilibrium do not guarantee the uniqueness of equilibrium in the current setting; and the following section will demonstrate that essentially no other general statements on the implications of ethical consumerism can be made in the current setting.

4 Implications for consumer behavior

An unfortunate feature of the results of Section 3 is that the informational requirements for computing ψ^i are very demanding. To decide what to buy, an ethical consumer i is asked to know the gradient of the aggregate Engel curve and the Jacobians of supply and implicit demand with respect to price (or, equivalently, the cross-price elasticity matrices of supply and implicit demand and the aggregate supply levels). One might therefore wonder whether it is possible to draw at least some incomplete conclusions about i's externality weights or optimal consumption basket directly from his preferences, without relying on much or any economic data.

Section 4.1 notes a narrow case in which i can easily infer ψ^i from $w^i(\cdot)$: $\psi^i = 0$ if (and only if) $\nabla w^i(s) \propto p$. Section 4.2, however, shows that essentially nothing else can be known about the relationship between $w^i(\cdot)$ and ψ^i from first principles. Section 4.3 details an implication of this point: that even if $w^i(\cdot)$ increases in every good, ψ^i may be strictly negative, and indeed so negative that i prefers not to exhaust his budget.

4.1 Prices and external preferences

Proposition 7. ***Externality weights zero if external preferences proportional to prices***

Let (p, ψ) be an RCESE.

a. If $\nabla w^i(s(p)) \propto p$, then $\psi^i = 0$.

b. If $\nabla w^i(s(p)) \not\propto p$ and $\text{Rank}(Ds(p)) = L - 1$, then $\psi^i \neq 0$.

Proof. Part (a) follows from definition (307) modified for the RCESE context—

$$\psi^i(p, \psi) \equiv (Ds(p)G(p, \psi))^T \nabla w^i(s(p)) \tag{323}$$

—and the fact that $p \cdot Ds(p) = \mathbb{0}_L^T \; \forall p$.

Part (b) follows from (323) and the fact that, if (p, ψ) is an RCESE, then $-Dz_{[\psi]}(p)$ and so also its generalized inverse $G(p, \psi)$ are of rank $L - 1$. Since the bottom row of $G(p, \psi)$ consists of zeroes, its left null space $\text{Null}(G(p, \psi)^T)$ is spanned by the unit row vector with a one in place L. The column space of $Ds(p)^T$ is orthogonal to its left null space, which is spanned by p. So if $\text{Rank}(Ds(p)) = L - 1$ and $\nabla w^i(s(p)) \not\propto p$, then $Ds(p)^T \nabla w^i(s(p)) \perp p$, so $Ds(p)^T \nabla w^i(s(p)) \notin \text{Null}(G(p, \psi)^T)$. □

In Proposition 4 and its proof, we noted trivially that if a consumer i is indifferent to supply levels ($w^i(\cdot) = 0$), he is also indifferent to potential purchases' impacts on supply levels ($\psi^i = 0$). Part (a) shows that this antecedent is stronger than necessary. A consumer may feel that the production of each good has desirable or undesirable consequences, but as long as he feels the significance of these consequences to be proportional on the current margin to the goods' prices, he puts no externality weight on any marginal purchases. He may support policies that would increase or decrease aggregate productive capacity, but he is indifferent to supply-shifts of the only sort that consumer decisions (at least in this model) can achieve: namely shifts within the hyperplane that contains $s(p)$ and is orthogonal to p.

Part (b) then shows that the $\nabla w^i(s(p)) \propto p$ condition cannot in general be weakened further. If a consumer feels that the the production of goods has desirable or undesirable consequences not proportional on the current margin to the goods' prices, then unless there is a direction in which relative prices can shift without affecting supply levels at all, he does put a nonzero externality weight on the consumption of some good.

Part (a) tells us that p is in the null space of $(Ds(p)G(p,\psi))^T$. This result has a dual: p is *not* in the column space.

Proposition 8. ***Externality weights zero if proportional to prices***
Let (p,ψ) be an RCESE. Then $\psi^i \propto p$ (if and) only if $\psi^i = 0$.

Proof. By construction, $\delta(p,\psi) \cdot G(p,\psi)^T = 0$. But $\delta(p,\psi) \not\perp p$, since by definition of RCESE $\frac{\partial}{\partial \pi}(p \cdot x_{[\psi]}(p,\pi)) = p \cdot \delta(p,\psi) > 0$. If $\psi^i(p,\psi) \propto p$ but $\neq 0$, left-multiplying both sides of (323) by $\delta(p,\psi)$ yields $\delta(p,\psi) \cdot \psi^i(p,\psi) = 0$, a contradiction. □

For interpretation, suppose $\psi^i \propto p$. Then any demand-shift dx^i that keeps i on

her budget constraint ($dx^i \perp p$) has on balance no externality weight ($\psi^i \cdot dx^i = 0$). Any dx^i can be decomposed into a weighted sum of a vector orthogonal to p and a vector proportional to δ, which changes her total expenditure but does not affect prices or supply. So if there is no externality weight associated with a demand-shift $dx^i \perp p$, there is no externality weight associated with any demand-shift.

Observe that, as long as $\chi^i_{[\psi]}(p)$ is interior and $v^i(\chi^i_{[\psi]}(p))$ is locally differentiable with respect to χ^i, the optimality of i's demand requires

$$\nabla_{\chi^i} v^i(\chi^i_{[\psi]}(p)) + \psi^i \propto p.$$

An implication of Proposition 8, therefore, is that if i does put some nonzero externality weight on the consumption of any good—e.g. if the conditions of Proposition 7b hold—then this in turn must somehow affect her consumption decisions, given the interiority and differentiability conditions.

4.2 Anything goes

Toward the end of Section 3, under "Existence more generally", we briefly observed how economic primitives, in particular consumption utility functions, can be chosen so that the economy admits a given profile of externality weights in equilibrium. This section extends that observation. For *any* relationship between ethical preferences $w^i(\cdot)$ and externality weights ψ^i except those ruled out in Section 4.1, a well-behaved economy can be constructed to match it.

Recall the Sonnenschein-Mantel-Debreu theorem. The following is a weakening of it sufficient to suggest the application to our case:

Given a price vector p and an $L \times L$ matrix A satisfying $Ap = \mathbb{0}_L$ and $p \cdot A = \mathbb{0}_L^T$, there

is an economy $\mathcal{E}$, with $v^i(\cdot)$ nondecreasing, locally nonsatiated, and strictly concave for all i, such that $z_{[\mathbb{0}_{L\times I}]}(p) = 0$ and $Dz_{[\mathbb{0}_{L\times I}]}(p) = A$.

In short, price-changes may induce arbitrary changes to excess demand around an equilibrium price vector.[16]

In the present context, the theorem has a dispiriting converse. Changes to excess demand—in particular, the changes an individual might make to her own demand function, holding others' fixed—may induce arbitrary changes to equilibrium prices, and thus to quantities supplied. It then follows that as long as prices p are not proportional to $\nabla w^i(s(p))$, the only constraint we can impose from first principles on i's equilibrium externality weights is that they too are not proportional to p (see Propositions 7a and 8).

Proposition 9. ***Anything goes***

Choose

1. *a profile of I $\mathcal{C}^1$ functions $\{w^i(\cdot)\}$ from $\mathbb{R}^L_{++}$ to $\mathbb{R}$, with $I \geq 2L+1$ and $L \geq 2$;*
2. *an L-vector $s \gg 0$;*
3. *an L-vector $p \gg 0$; and*
4. *an $L \times L$ matrix M with $p \in \text{Null}(M)$ and $p \notin \text{Col}(M)$,*

and define ψ by $\psi^i = M\nabla w^i(s)\ \forall i$.

Then there is an environment E and a profile of increasing, strictly concave consumption utility functions $\{v^i(\cdot)\}$ such that $s(p) = s$ and (p, ψ) is an RCESE of economy $(E, \{v^i(\cdot) + w^i(\cdot)\})$.

[16]The theorem is therefore sometimes known as the "anything goes" theorem. The weakened statement of it above is adapted from Mas-Colell et al. (1995), Proposition 17.E.2. and Mantel (1974).

Proof. See Appendix B.6. □

The example of Section 3.1 hinted at Proposition 9, in that it featured an equilibrium externality vector with non-positive entries ($\psi^i = (-1/3, 0)^T$) for an individual i with $\nabla w^i \gg 0$. To illustrate the the strength of the proposition more fully, consider the case in which $w^i(\cdot) \equiv w(\cdot)$ is the same for all i and $M\nabla w(s) = -\nabla w(s)$. Suppose for example that $p = (1, 1)^T$, $\nabla w(s) = (1, -2)^T$, and

$$M = \frac{1}{3}\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix},$$

so that $\psi^i = (-1, 2)^T \ \forall i$. Conditions can be constructed under which (p, ψ) is an RCESE. Though every individual agrees that it would be preferable to produce more of good 1 and less of good 2, every individual's ethical considerations weigh against buying the former and in favor of buying the latter.

4.3 Satiation

To illustrate the strength of Proposition 9 perhaps even more fully, we will now construct an economy and an RCESE in which, for some i, $\nabla w^i \gg 0$ but $\psi^i \ll 0$—and in which the negative externality weight for i of any marginal purchase is large enough that, though $\nabla v^i \gg 0$ as well, i prefers not to exhaust her budget. We will then discuss what drives this result.

Example: "no ethical consumption under capitalism"

Consider the following economy $\mathcal{E}$ with $I = 2$ individuals and $L = 2$ goods:

$$
\begin{aligned}
e^1 &= (1,1)^T \quad \forall i; \\
\theta^1 &= 0,\ \theta^2 = 1; \\
v^1(x^1) &= \frac{4}{3}\ln(x_1^1) - \frac{2}{15}\exp(-x_2^1), \qquad (324) \\
v^2(x^2) &= \ln\left(x_1^2 - \frac{1}{2}\right) - 2\ln(3 - x_2^2) \ \text{ around } x^2 = (1,2)^T, \qquad (325) \\
&\qquad \text{locally nonsatiated and globally strictly quasiconcave;} \\
w^1(s) &= s_1 + 2s_2, \\
w^2(s) &= 0;
\end{aligned}
$$

and whose production possibilities frontier is given by

$$y_2 = 2 - (1 + y_1)^2/4,$$

so that

$$s_1(p) = 2p_1/p_2, \quad s_2(p) = 3 - (p_1/p_2)^2. \qquad (326)$$

We will show that $(\bar{p}, \bar{\psi})$ with $\bar{p} = (1,1)^T$, $\bar{\psi}^1 = (-4/3, -2/15)^T$, $\bar{\psi}^2 = 0$ is an RCESE of $\mathcal{E}$, with an associated profit level $\bar{\pi} = 1$. At this RCESE, individual 1 will be seen to spend only half her budget.

Given (324), the value of x^1 that globally maximizes $\tilde{u}^1_{[\bar{\psi}^1]}(x^1)$ is $(1,0)^T$. This basket is more than affordable for individual 1 at $(\bar{p}, \bar{\pi})$: its cost equals 1 and individual 1's budget equals 2. So $x^1_{[\bar{\psi}^1]}(p, \pi) = (1,0)^T$ around $(\bar{p}, \bar{\pi})$, locally independent

of p and (because $\theta^1 = 0$) globally independent of π.

Given (326), $s(\bar{p}) = e = (2, 2)^T$ and

$$Ds(\bar{p}) = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}. \tag{327}$$

Given (325), $x^2_{[\mathbb{0}_2]}(p, \pi)$ is globally defined, by strict quasiconcavity; $x^2_{[\mathbb{0}_2]}(\bar{p}, \bar{\pi}) = (1, 2)^T$; and, fixing $p_2 = 1$, so that 2's budget equals $p_1 + 1 + \pi$,

$$x^2_{1[\mathbb{0}_2]}(p, \pi) = \frac{2 - \pi}{p_1}, \qquad x^2_{2[\mathbb{0}_2]}(p, \pi) = p_1 + 2\pi - 1 \tag{328}$$

around $(\bar{p}, \bar{\pi})$. Good 1 is thus locally inferior, and good 2 is locally normal, for individual 2. Recalling that $x^1(\cdot)$ is independent of π,

$$\delta(\bar{p}, \bar{\psi}) = (1, 2)^T. \tag{329}$$

Since $p \cdot x^2_{[\mathbb{0}_2]}(p, \pi)$ strictly increases without bound in π, by local nonsatiation, $\pi(p, \{x_{[\bar{\psi}]}(\cdot)\})$—the unique value of π setting

$$\bar{p} \cdot x_{[\bar{\psi}]}(\bar{p}, \pi) = p \cdot s(\bar{p})$$

—is defined for all p. Around $\bar{p}$, still fixing $p_2 = 1$, we have

$$\pi(p, \{x_{[\bar{\psi}]}(\cdot)\}) = \frac{2p_1^2 - 6p_1 + 5}{2 - p_1}. \tag{330}$$

Substituting 1 for p_1, we have $\bar{\pi} \equiv \pi(\bar{p}, \{x_{[\bar{\psi}]}(\cdot)\}) = 1$.

Substituting (330) for π into (328) gives

$$\chi^2_{1[\bar{\psi}]}(p) = \frac{2p_1^2 - 4p_1 + 1}{p_1^2 - 2p_1}, \qquad \chi^2_{2[\bar{\psi}]}(p) = \frac{3p_1^2 - 9p_1 - 8}{2 - p_1}. \tag{331}$$

Substituting 1 for p_1, we have $\chi^2_{[\bar{\psi}]}(\bar{p}) = (1,2)^T$. Since aggregate demand $(1,0)+(1,2)$ equals supply $(2,2)$, this confirms that $\bar{p}$ is a WES of $(\mathcal{E}, x_{[\bar{\psi}]}(\cdot))$.

Differentiating (331) with respect to p_1 and using the fact that $\chi^2(\cdot)$ must be h.o.d. 0 gives

$$D\chi^2_{[\bar{\psi}]}(\bar{p}) = \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix}. \tag{332}$$

Then (trivially) adding (332) to $D\chi^1_{[\bar{\psi}]}(\bar{p}) = \mathbb{0}_{2\times 2}$ to get $D\chi_{[\bar{\psi}]}(\bar{p})$; subtracting this from (327) to get $-Dz_{[\bar{\psi}]}(\bar{p})$; taking the appropriate generalized inverse compatible with (329) to get $G(\bar{p}, \bar{\psi})$; and left-multiplying by (327), transposing, and right-multiplying by $\nabla w^1 = (1,2)^T$ confirms that $\psi(\bar{p}, \bar{\psi}) = \bar{\psi}$.

Four features of the economic setup drive the result above. Around the proposed equilibrium,

i. Individual 1 prefers to shift production from good 1 to good 2.

ii. Good 1 is inferior for individual 2.

iii. For individual 1, the marginal consumption utility of good 1 is significantly higher than the marginal consumption utility of good 2.

iv. For individual 1, the marginal consumption utility of each good is low, relative to the external utility achievable by shifting supply.

By (i) and (ii), individual 1 shifts supply in a preferred direction when she decreases her spending. This effectively transfers income to individual 2, which he uses

to shift supply toward good 2.

Individual 1 can increase her spending without impacting supply. However, this requires increasing her spending along the gradient $\bar{\delta}$: buying one fewer unit of good 1 and two more units of good 2 with each additional dollar spent instead of transferred to 2, so that the individuals' consumption baskets shift in opposite directions. By (iii), individual 1 prefers not to increase her spending in this direction. More generally, any spending increase by individual 1 with a weakly desirable supply impact would require a large enough decrease to x_1^1 per unit of increase to x_2^1 as to lower her consumption utility.

Finally by (iv), the negative externality weight associated with any spending increase by individual 1 that would increase her consumption utility is high enough that no spending increase is desirable.

Non-satiation conditions

Though non-satiation cannot be assumed to hold in general, it can under some circumstances. One such circumstance is as follows.

Proposition 10. ***Non-satiation in RCESE given no aggregate inferior goods***

Given an RCESE $(\bar{p}, \bar{\psi})$, suppose $\delta(\bar{p}, \bar{\psi}) \geq 0$. Then $\bar{\psi}^i \not\ll 0 \ \forall i$. If $v^i(\cdot)$ is locally nonsatiated for all i, then all i exhaust their budgets.

Proof. For all (p, ψ), we have $G(p, \psi)\delta(p, \psi) = 0$, and thus

$$\psi^i(p, \psi) \cdot \delta(p, \psi) = \nabla w^i(s(p)) \cdot Ds(p)G(p, \psi)\delta(p, \psi) = 0 \ \ \forall i. \tag{333}$$

Since $\delta(\bar{p}, \bar{\psi}) \geq 0$, it follows that $\psi^i(\bar{p}, \bar{\psi}) \not\ll 0$. For each i, because $v^i(\cdot)$ is locally

nonsatiated, it follows that there is a good in which i's consumption utility is non-satiated in equilibrium. Therefore all individuals exhaust their budgets. □

Proposition 10 tells us that as long as $\delta \geq 0$—that is, as long as each good ℓ is not "inferior in aggregate", in that x_ℓ is not locally decreasing in the profit rate—then each i can infer that it will be optimal for her to exhaust her budget. As a result, i does not need to choose her demands by constructing G, and then ψ^i, using known δ. Instead, i can recognize that *any* generalized inverse $\tilde{G}$ of $-Dz$ satisfies $-Dz(\bar{p})\tilde{G}dx^i = dx^i$ for $dx^i \perp \bar{p}$, and thus that $\tilde{G}$ captures the price impacts,[17] and pins down the supply impacts, of demand choices among baskets that exhaust i's budget. Therefore, in this setting,

$$x_{[\tilde{\psi}^i]}(p,\pi) = \underset{x^i}{\operatorname{argmax}}\, \tilde{u}^i(x^i, \tilde{\psi}^i) \;\Big|\; p \cdot x^i = p \cdot e^i + \theta^i \pi \tag{334}$$

for any $\tilde{\psi}^i = \left(Ds(\bar{p})\tilde{G}\right)^T \nabla w^i(s(\bar{p}))$, where $\tilde{G}$ is an arbitrary generalized inverse of $-Dz(\bar{p})$.

Corollary 10.1. ***Non-satiation in RCESE given separability of consumption utility***

Suppose that, for all i, $v^i(\cdot)$ is additively separable and locally nonsatiated. Then, in any RCESE $(\bar{p}, \bar{\psi})$, $\bar{\psi}^i \not\ll 0\ \forall i$, and all individuals exhaust their budgets.

Proof. The additive separability of the $v^i(\cdot)$ and thus of the $\tilde{u}^i(\cdot)$ implies that, at *any* (p, ψ), no good is inferior for any i. Therefore no good is inferior in aggregate. Recalling that $\delta(\bar{p}, \bar{\psi})$ must be defined if $(\bar{p}, \bar{\psi})$ is an RCESE, we have $\delta(\bar{p}, \bar{\psi}) \geq 0$. The result then follows from Proposition 10. □

[17] Up to rescaling. Without loss of generality, we can restrict ourselves to price impacts with $dp_L = 0$ by imposing that the bottom row of $\tilde{G}$ equal $\mathbb{0}_L^T$.

Proposition 10 and its corollary thus demonstrate that, under some conditions on the economic primitives, a consumer with external preferences does not need to know the gradient of the aggregate Engel curve in order to compute her optimal demands. It is possible that further inquiry will likewise identify reasonable conditions under which she can dispense with some of the price substitution variables. There thus appears to be at least some grounds for hope that, though theory alone offers no guidance to the ethical consumer, in some circumstances the project of ethical consumerism is feasible.

5 Conclusion

Spending decisions are sometimes motivated, in whole or part, by a desire to make the world a better place by changing quantities supplied. In the simplest case, people engage in philanthropy. When individuals act on their preferences over supply—what Andreoni (1990) calls altruism, and what we have called ethical or external preferences—they engage in a public good game.

A conventional consumer can choose her demands on the basis of her budget and the price vector, without consulting any further economic data. This is not true of someone with external preferences. Conventional models of public good games partially acknowledge this. When deciding how to allocate his budget across public goods, an individual philanthropist's best response to other philanthropists' spending decisions is recognized to depend not only on his own budget and the goods' prices but also on the aggregate quantities being provided by others. However, when prices and budgets are assumed to be exogenous to the philanthropist's spending decisions, he is spared from having to consider the income and price substitution effects of other market participants. Like the conventional consumer, he can act as if the

world is a vending machine that linearly transforms currency into goods. Unless the production possibility set is effectively a simplex, so that relative prices are in fact exogenous, such analyses have been incomplete.

This paper and Kaufmann et al. (2024) have taken steps toward completing them by embedding individuals with external preferences in an explicit production economy. This has for the first time allowed a characterization of how optimal spending behavior by an individual with external preferences depends on the broader array of relevant economic variables. In both papers, this characterization reveals that important behaviors that economic agents in a competitive setting are typically assumed to exhibit—in particular price-taking—cannot be assumed when the agents have external preferences.

The present paper further demonstrates that important behaviors exhibited in a simple partial equilibrium setting (or a conventional public good provision setting) cannot be assumed in a general equilibrium setting. One cannot even straightforwardly assume budget-exhaustion, nor a sign-preserving relationship between the preference for increasing the quantity of a good and the willingness to pay for a unit of it in the marketplace. A lesson for those with external preferences is that to determine which purchases are optimal, or even desirable, one must gather and process much more economic data than has hitherto been assumed necessary or is typically done in practice: data almost analogous to what one would need to plan an economy. If one draws the pessimistic conclusion that predicting the impacts of small individual spending decisions on aggregate quantities is typically infeasible (and if one trusts that the effects of at least some kinds of policy are more often predictable), perhaps a broad further lesson is that resources spent on making the world a better place through the market are typically better directed to influencing

policy.

Then by studying economies with multiple ethical consumers, this paper and that of Kaufmann et al. independently identify the natural equilibrium concept that emerges from their interaction. This lets us confirm that the nonstandard behavior which individuals with external preferences optimally adopt can coherently be adopted by many, even all, at once.

If we maintain the hope that it is sometimes feasible to account for the general equilibrium effects of our spending decisions, a natural next step is to identify conditions under which the these effects can be approximated with less data than would be required to calculate the expressions given here. Such work would be valuable to individual ethical consumers, donors, and philanthropists, and may make the associated behaviors more appealing by putting estimates of their consequences on surer footing. It may also shed light on the implications of philanthropy and ethical consumerism for social welfare and optimal policy in various settings, as Kaufmann et al. are able to do in their setting.

Complicating matters, on the other hand, further research is necessary to understand optimal behavior by individuals with preferences not directly over aggregate quantities of various goods (as most clearly in the case of concern for animal welfare) but over other individuals' consumption utility levels. Preferences of the latter kind can generate nonzero externality weights associated with a given good purchase even in an exchange economy, since, as emphasized by Wilkinson (2022), purchases can affect the distribution of income.

Further research is also necessary to understand the impacts of marketplace decisions through channels other than that considered here.

For instance, though we have assumed a single, profit-maximizing firm, individ-

uals with external preferences often aim to direct capital away from firms whose production processes they see as having negative externalities (via divestment) or toward those with positive externalities (via impact investing). These efforts obey a logic similar to the equilibrium displacement logic we have seen in the case of consumer spending.[18] It would evidently be straightforward, therefore, to introduce a model of optimal investment by agents with external preferences along the lines of the model of ethical consumerism introduced here. A unified model of the economic implications of external preferences in both domains may also identify important interactions between them.

Moreover, we have studied a static economy, but the most significant consequences of spending decisions may arise only in a dynamic setting. For instance, increases in demand for a good today may result in better technology for producing the good in future periods, either directly through "learning by doing" or by motivating profit-seeking firms to research improved methods for producing goods they expect to remain in high demand. This suggests the value of embedding a model of ethical consumerism in a model of economic growth.

[18]Because investors value diversification, the supply of capital each firm or industry receives is not perfectly elastic in the interest rate it offers. An individual investor's decision to increase or decrease his own investment in it therefore affects its size in equilibrium, even in an approximately competitive environment.

This point is made clearly by Christiano (2019), but to my knowledge, the financial economics literature on impact investing to date has neglected it. Instead either

a) investors are assumed to be infinitesimal and thus fully price-taking, and invest in ethical firms offering below-market returns only because they have an intrinsic preference for doing so (somewhat analogously to Dufwenberg et al. (2011), discussed in Section 1 of this paper: see e.g. Pastor et al. (2021) or Roth Tran (2019), beginning of Section IV);

b) an investor with strictly positive market power is considered (see e.g. Roth Tran (2019), end of Section IV, or Roth (2021)); or

c) firms are assumed to face a perfectly elastic supply of capital at a fixed interest rate, so that ethical investors can have an impact only by investing in projects offering below-market returns which otherwise would have received no investment and which therefore crowd out no investment (Green and Roth, 2024).

Finally, we have of course assumed a smooth, frictionless, and perfectly competitive environment. There are surely intuitive grounds for believing that frictions—e.g. the costs of altering production capacity for some good—in some sense do not alter the logic of continuous equilibrium displacement "on average" (see e.g. Isaacs et al. (2024)). Some (e.g. Budolfson (2019)) have disputed this, however, arguing informally that they imply that small consumers may have approximately no impact on supply even in expectation. This debate can only be adjudicated by carefully studying the frictions in question and their implications for the path between supply and marginal consumer behavior. If there is any lesson in the limited modeling done so far on questions of strategic equilibrium displacement, it is that applying economic intuitions developed for other purposes can lead us astray. Effects of a purchase that might at first seem insignificant, such as infinitesimal price-changes or income effects induced in others, can be of primary importance to the ethical consumer.

Appendices

A Supplemental materials

A.1 The normal form game

Players, strategies, and outcomes

The strategic interaction we study is fully identified by the economy $\mathcal{E}$. Up to one technical complication, detailed below, it may be defined as a normal form game as follows.

- Players: There are $I + 1$ players, I individuals and nature.

- Strategy sets:
  - i's strategy set is $\mathbb{A}^i(\mathcal{E})$.
  - Nature chooses a continuous function $f(\cdot) : \mathbb{R}^{L-1} \to \mathbb{R}$.
- Outcomes: Denote the set of NRWES of $(\mathcal{E}, \{x^i(\cdot)\})$ by R. If

$$\hat{\bar{p}} \equiv \underset{\hat{p} \in R}{\operatorname{argmax}} f(\hat{p})$$

  exists and is unique, $\hat{\bar{p}}$ the NRWES that obtains. All economic outcomes—supply, profits, individual demands, and thus individuals' utility levels—are determined by $\hat{\bar{p}}$, given $(\mathcal{E}, \{x^i(\cdot)\})$. Nature is indifferent among outcomes.

Because we have not defined the outcome of all strategy profiles, the game defined above is a normal form game with incomplete preferences. Following e.g. Bade (2005), a strategy profile of such a game is a Nash equilibrium if the corresponding utility profile $\{\bar{u}^i\}$ is defined and if, for each player i, no deviation yields a strategy profile at which u^i is defined and greater than $\bar{u}^i$.

In particular, we have not defined the outcome of a strategy profile in the event that there is no RWES of $(\mathcal{E}, \{x^i(\cdot)\})$. This is unnecessary because CESE, and the equilibrium refinement we consider in Section 3.2 (RCESE), presuppose strategy profiles compatible with an RWES.

We have also not defined the outcome of a strategy profile in the event that there are multiple NRWES and none uniquely maximizes $f(\cdot)$. The difficulty in this case is subtle, as detailed in the subsection just below, and is perhaps the reason that to my knowledge no existing literature on the strategic foundations of general equilibrium has presented the strategic interaction under study as a normal form game.

Selection of economic equilibrium

Suppose that $\hat{p}$ is an NRWES compatible with $(\mathcal{E}, \{x^i(\cdot)\})$ and that $\mathcal{E}$ is large relative to $\{x^i(\cdot)\}$. By Proposition 2, there is a unique NRWES near $\hat{p}$ compatible with the demand function profile $\{\tilde{x}^i(\cdot)\}$ in which a single i deviates from $\{x^i(\cdot)\}$.

Since a demand function profile may be compatible with multiple NRWES, we would like to say that nature identifies the *approximate* location of the NRWES, so that deviations by any individual produce only slight shifts to equilibrium prices. But there is no continuous, complete ordering on $\mathbb{R}^k$ for $k > 1$. If $L > 2$, therefore, there is no way to ensure that the NRWES nature chooses among those compatible with the demand function profile with i's deviation is always that NRWES near $\hat{p}$. Nature may be, as it were, on the fence between NRWES $\hat{p}$ and NRWES $\hat{p}'$, in which case a slight change to an individual's demand can discontinuously shift equilibrium prices.

To illustrate the difficulty, suppose that $L = 3$ and that nature orders NRWES lexicographically, choosing that with the highest value of p_1 if this is unique and that with the highest value of p_2 if multiple equally have the highest value of p_1. The lexicographic ordering is strict; there are no ties. But a chosen NRWES is then not globally continuous in the demand function profile. If the NRWES compatible with $\{x^i(\cdot)\}$ are $\hat{p} = (1, 2)$ and $\hat{p}' = (1, 1)$, $\hat{p}$ is chosen; but if the NRWES compatible with $\{\tilde{x}^i(\cdot)\}$ are $\tilde{\hat{p}} = (1, 2)$ and $\tilde{\hat{p}}' = (1.01, 1)$, $\tilde{\hat{p}}'$ is chosen.

Under appropriate assumptions on $f(\cdot)$, however, we can ensure that, for any $m > 0$, almost all sets of m normalized price vectors have exactly one that maximizes $f(\cdot)$. For this reason, and to make our analysis as similar as possible to the existing literatures on equilibrium displacement and on the strategic foundations of general equilibrium, we ignore this complication in the body of the paper. Implicitly, we

assume that whenever there is at least one NRWES of $(\mathcal{E}, \{x^i(\cdot)\})$, nature chooses one $\hat{p}$ among them; and that nature's strategy, fortuitously, is such that if any i deviates to a different demand function, the chosen NRWES is always that near to $\hat{p}$.

Relationship between CESE and Nash equilibrium

Call strategy profile $(\{x^i(\cdot)\}, f(\cdot))$ a CESE* of economy $\mathcal{E}$ if the implied NRWES $\hat{\hat{p}}$ is defined and $(\hat{\hat{p}}, \{x^i(\cdot)\})$ is a CESE.

CESE* is slightly more restrictive than CESE, in that it presupposes not only that there is at least one RWES of $(\mathcal{E}, \{x^i(\cdot)\})$ but also that a unique NRWES is identified by the economy and strategy profile $(\mathcal{E}, \{x^i(\cdot)\}, f(\cdot))$.

CESE* is a weakening of pure-strategy Nash equilibrium. Whereas Nash equilibrium maintains that no player can benefit by deviating, CESE* maintains only that the benefits to deviating a player would be able to achieve in an n-replicated economy and demand function profile fall to zero as $n \to \infty$.

A.2 Externality weights in elasticity terms

For convenience in applications we can express (307) in elasticity terms, as long as $s(p) \gg 0$.

Given prices $p \gg 0$, let $\sigma(p)$ denote the matrix of cross-price elasticities of supply around prices p and $\varepsilon(p, x(\cdot))$ denote the matrix of cross-price elasticities of implicit demand around p given aggregate demand function $x(\cdot)$. Given an L-vector b, let $\text{Diag}(b)$ denote the $L \times L$ diagonal matrix with the entries of b on its diagonal. Finally

let

$$\phi(p, x(\cdot)) \equiv \operatorname{Diag}(p)^{-1}\, G(p, x(\cdot))\, \operatorname{Diag}(s(p)) \tag{335}$$

denote the generalized inverse of $(\sigma(p) - \varepsilon(p, x(\cdot)))$ with

$$\phi(\bar{p}, \bar{x}(\cdot))\, \operatorname{Diag}(s(p))^{-1}\, \delta(p, x(\cdot)) = 0$$

and whose bottom row consists of zeroes. Such a generalized inverse within the hyperplane perpendicular to p exists, is unique, and is equal to expression (335) because for $dx^i \perp p$ we have

$$\begin{aligned} & -Dz(p)\, G(p, x(\cdot))\, dx^i = dx^i \\ \implies & \operatorname{Diag}(s(p))\, (\sigma(p) - \varepsilon(p, x(\cdot)))\, \operatorname{Diag}(p)^{-1}\, G(p, x(\cdot))\, dx^i = dx^i \\ \implies & (\sigma(p) - \varepsilon(p, x(\cdot)))\, \phi(p, x(\cdot))\, \operatorname{Diag}(s(p))^{-1}\, dx^i = \operatorname{Diag}(s(p))^{-1}\, dx^i. \end{aligned}$$

(Observe that given $s(p) \gg 0$, $dx^i \perp p$ iff $\operatorname{Diag}(s(p))\, dx^i \perp p$.)

Substituting

$$\begin{aligned} G(p, x(\cdot)) &= \operatorname{Diag}(p)\, \phi(p, x(\cdot))\, \operatorname{Diag}(s(p))^{-1}, \\ Ds(p) &= \operatorname{Diag}(s(p))\, \sigma(p)\, \operatorname{Diag}(p)^{-1} \end{aligned}$$

into (307) yields

$$\begin{aligned} \psi^i(p, x(\cdot)) &= \big(\operatorname{Diag}(s(p))\, \sigma(p)\, \phi(p, x(\cdot))\, \operatorname{Diag}(s(p))^{-1}\big)^T \nabla w^i(s(p)) \\ &= \big(\sigma(p)\, \phi(p, x(\cdot))\big)^T \nabla w^i(s(p)). \end{aligned}$$

B Proofs

B.1 Proof of Proposition 1

Given economy $\mathcal{E}$ and aggregate demand function $x(\cdot) \in \mathbb{A}(\mathcal{E})$, consider $(\mathcal{E}^{(n)}, x^{(n)}(\cdot))$. Excess demand in the replicated setting equals

$$z^{(n)}(p) = nx(p, \pi^{(n)}(p, x^{(n)}(\cdot))/n) - ns(p), \tag{336}$$

where aggregate supply in the replicated economy equals $ns(p)$ because the aggregate endowment and production function have both been multiplied by n, and where

$$\begin{aligned} \pi^{(n)}(p, x^{(n)}(\cdot)) &\equiv \pi : x^{(n)}(p, \pi) = ns(p) && (337) \\ &= \pi : x(p, \pi/n) = s(p) \\ &= n\big(\pi : x(p, \pi) = s(p)\big) \\ &= n\pi(p, x(\cdot)). && (338) \end{aligned}$$

Substituting (338) into (336), we have $z^{(n)}(p) = nz(p)$. It follows immediately that p is a [R]WES of $(\mathcal{E}, x(\cdot))$ iff it is a [R]WES of $(\mathcal{E}^{(n)}, x^{(n)}(\cdot))$, for any $n \geq 1$.

Define

$$Z(p, \pi, x(\cdot)) \equiv p \cdot x(p, \pi) - p \cdot s(p).$$

Because $x(p, \pi)$ is h.o.d. 0 in (p, π) by admissibility and $s(p)$ is h.o.d. 0 in p, $Z(p, \pi, x(\cdot))$ is h.o.d. 0 in (p, π). In particular, if $Z(p, \pi, x(\cdot)) = 0$, $Z(kp, k\pi, x(\cdot)) = 0 \; \forall k$. It follows that $\pi(p, x(\cdot))$ is is h.o.d. 1 in p.

Thus

$$\begin{aligned} z(kp) &= x(kp, \pi(kp, x(\cdot))) - s(kp) \\ &= x(kp, k\pi(p, x(\cdot))) - s(kp) \\ &= x(p, \pi(p, x(\cdot))) - s(p) \\ &= z(p); \end{aligned}$$

$z(p)$ is h.o.d. 0 in p. So any positive rescaling of $\bar{p}$ is also a WES. In particular, $\check{\hat{p}}$ is a WES. Thus $\hat{\bar{p}}$ is an NWES.

Conversely, if $\hat{\bar{p}}$ is an NWES, any positive rescaling of $\check{\hat{p}}$, including $\bar{p}$, is a WES.

B.2 Proof of Proposition 2

Let $\bar{p}$ be a regular WES of economy $\mathcal{E}$ and locally regular demand function profile $\{\bar{x}^i(\cdot)\} \in \mathbb{A}^i(\mathcal{E})$. Choose $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$ that is $\mathcal{C}^1$ around $(\bar{p}, \bar{\pi})$.

Let

$$h(\pi, \hat{p}, \alpha) \equiv \check{\hat{p}} \cdot \big(\bar{x}(\check{\hat{p}}, \pi) + \alpha\big(x^i(\check{\hat{p}}, \pi) - \bar{x}^i(\check{\hat{p}}, \pi)\big) - s(\check{\hat{p}})\big). \tag{339}$$

Because $s(\cdot)$, $x^i(\cdot)$, and each $\bar{x}^i(\cdot)$ are $\mathcal{C}^1$ around $(\check{\hat{\bar{p}}}, \bar{\pi}/p_L)$, and because the composition of $\mathcal{C}^1$ functions is $\mathcal{C}^1$, $h(\cdot)$ is $\mathcal{C}^1$ around $(\bar{\pi}/p_L, \hat{\bar{p}}, \alpha)$ for all α.

By construction of the profit function, $h(\bar{\pi}, \hat{\bar{p}}, 0) = 0$. Also, by the local regularity of $\{\bar{x}^i(\cdot)\}$, $\partial h/\partial \pi$ is nonzero around $(\bar{\pi}, \hat{\bar{p}}, 0)$. Thus, by the IFT, there is a unique, $\mathcal{C}^1$ function, which we will denote $\underline{\hat{\pi}}(\hat{p}, \alpha)$, such that $\underline{\hat{\pi}}(\hat{\bar{p}}, 0) = \pi(\check{\hat{\bar{p}}}, \bar{x}(\cdot)) \equiv \bar{\pi}$ and $h(\underline{\hat{\pi}}(\hat{p}, \alpha), \hat{p}, \alpha) = 0$ for all $(\hat{p}, \alpha)$ near $(\hat{\bar{p}}, 0)$.

Now, let

$$\underline{\hat{z}}(\hat{p},\alpha) \equiv \mathcal{I}\big[\bar{x}(\check{\hat{p}},\underline{\hat{\pi}}(\hat{p},\alpha)) + \alpha\big(x^i(\check{\hat{p}},\underline{\hat{\pi}}(\hat{p},\alpha)) - \bar{x}^i(\check{\hat{p}},\underline{\hat{\pi}}(\hat{p},\alpha))\big) - s(\check{\hat{p}})\big]. \tag{340}$$

Because $s(\cdot)$, $x^i(\cdot)$, and each $\bar{x}^i(\cdot)$ are locally $\mathcal{C}^1$, and $\underline{\hat{\pi}}(\cdot)$ is $\mathcal{C}^1$, it follows that $\underline{\hat{z}}(\cdot)$ is locally $\mathcal{C}^1$ around $(\hat{\bar{p}}, 0)$.

Because $\bar{p}$ is a WES of $(\mathcal{E}, x(\cdot))$, $\underline{\hat{z}}(\hat{\bar{p}}, 0) = 0$. Also, because $\bar{p}$ is a regular WES, the Jacobian of $\underline{\hat{z}}(\cdot)$ with respect to prices is nonsingular at $(\hat{\bar{p}}, 0)$. Thus, by the IFT, there exist $\epsilon_1, \epsilon_2 > 0$ such that, for every $\alpha \in (-\epsilon_1, \epsilon_1)$, there is a unique $\hat{p} \in \mathcal{N}_{\epsilon_2}(\hat{\bar{p}})$ such that $\underline{\hat{z}}(\hat{p}, \alpha) = 0$.[19] Furthermore, defining

$$\hat{g}(\alpha) \equiv \hat{p} : \underline{\hat{z}}(\hat{p}, \alpha) = 0, \quad \alpha \in (-\epsilon_1, \epsilon_1), \tag{341}$$

$\hat{g}(\cdot)$ is $\mathcal{C}^1$.

Thus

$$\underline{\hat{z}}(\hat{g}(1/n), 1/n) = 0 \ \ \forall n \geq \underline{n} \equiv \lfloor 1/\epsilon_1 \rfloor + 1. \tag{342}$$

It follows that $\hat{g}(1/n)$ is a normalized WES, and $(\hat{g}(1/n), 1)$ is a WES, of economy $\mathcal{E}$ and aggregate demand function

$$x(\cdot) \equiv \bar{x}(\cdot) + (x^i(\cdot) - \bar{x}^i(\cdot))/n. \tag{343}$$

It also follows that $\hat{g}(1/n)$ is the unique normalized WES in $\mathcal{N}_{\epsilon_2}(\hat{\bar{p}})$.

Finally, because $\underline{\hat{z}}(\cdot)$ is locally $\mathcal{C}^1$, and because the determinant of its Jacobian with

[19] $\mathcal{N}_\epsilon(b)$ denotes the open neighborhood of radius ϵ around vector b.

respect to prices is nonzero at $(\hat{\bar{p}}, 0)$, there is an $\epsilon_3 > 0$ such that, for all $(\hat{p}, \alpha) \in \mathcal{N}_{\epsilon_3}((\hat{\bar{p}}, 0))$, the determinant of its Jacobian with respect to prices is nonzero at $(\hat{p}, \alpha)$. Therefore, as long as we pick ϵ_1 and ϵ_2 small enough that

$$\sqrt{\epsilon_1^2 + \epsilon_2^2} \leq \epsilon_3, \tag{344}$$

we can guarantee that, as long as $n \geq \lfloor 1/\epsilon_1 \rfloor + 1$, $\hat{g}(1/n)$ is not only a WES but a regular WES of $(\mathcal{E}, x(\cdot))$.

It then follows from Proposition 1 and the definition of n-replicated utility that, as long as $n \geq \lfloor 1/\epsilon_1 \rfloor + 1$, $\hat{g}(1/n)$ is a regular WES, and the unique normalized WES in $\mathcal{N}_{\epsilon_2}(\hat{\bar{p}})$, of $(\mathcal{E}^{(n)}, \bar{x}^{-i(n)}(\cdot) + x^{i(n)}(\cdot))$.

B.3 Proof of Proposition 3

Let $\bar{p}$ be a WES of economy $\mathcal{E}$ and locally regular demand function profile $\{\bar{x}^i(\cdot)\} \in \{\mathbb{A}^i(\mathcal{E})\}$.

Let n be large enough that $\mathcal{E}^{(n)}$ is large with respect to $\{\bar{x}^i(\cdot)\}^{(n)}$ and $\hat{\bar{p}}$. Then consider the change in $u_{\bar{p}}^{i(n)}$ that i achieves by deviating to locally $\mathcal{C}^1$ demand function $x^i(\cdot) \in \mathbb{A}^i(\mathcal{E})$:

$$\begin{aligned} &v^i\Big(x^{i(n)}\Big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big), \pi^{(n)}\big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big), \bar{x}^{-i(n)}(\cdot) + x^{i(n)}(\cdot)\big)\Big)\Big) + nw^i\Big(s\Big(\check{\hat{p}}_{\bar{p}}^{(n)}\big(x^i(\cdot)\big)\Big)\Big) \\ &- v^i\big(\bar{x}^{i(n)}(\bar{p}, \pi^{(n)}(\bar{p}, \bar{x}^{(n)}(\cdot)))\big) - nw^i\big(s(\bar{p})\big). \end{aligned} \tag{345}$$

In steps, we will take the limit of (345) as $n \to \infty$ and determine when the expression is nonpositive for any admissible choice of $x^i(\cdot)$.

First, by definitions (302) and (303), (345) equals

$$v^i\big(x^i\big(\check{\hat{p}}^{(n)}_{\bar{p}}\big(x^i(\cdot)\big), \pi\big(\check{\hat{p}}^{(n)}_{\bar{p}}\big(x^i(\cdot)\big), \bar{x}(\cdot) + (x^i(\cdot) - \bar{x}^i(\cdot))/n\big)\big)\big) - v^i\big(\bar{x}^i(\bar{p}, \bar{\pi})\big)$$
$$+nw^i\big(s\big(\check{\hat{p}}^{(n)}_{\bar{p}}\big(x^i(\cdot)\big)\big)\big) - nw^i\big(s(\bar{p})\big). \tag{346}$$

By Proposition 2 and the definition of $\hat{p}^{(n)}_{\bar{p}}(\cdot)$, we have

$$\check{\hat{p}}^{(n)}_{\bar{p}}(x^i(\cdot)) = g(1/n) \equiv (\hat{g}(1/n), 1), \tag{347}$$

where $\hat{g}(\cdot)$ is defined as in (341). Because $g(\alpha)$ is continuous and equals $\check{\hat{p}}$ at $\alpha = 0$, and because $x^i(p, \pi)$ is h.o.d. 0 and $\pi(p, x(\cdot))$ is h.o.d. 1 in p, the limit as $n \to \infty$ of the first term of (346) equals

$$v^i(x^i(\bar{p}, \bar{\pi})).$$

Substituting (347) into the third term of (346), and replacing n with $1/\alpha$ (where $\alpha \equiv 1/n$), the third and fourth terms equal

$$\frac{w^i\big(s(g(\alpha))\big) - w^i\big(s(\bar{p})\big)}{\alpha}. \tag{348}$$

Because $w^i(\cdot)$, $s(\cdot)$, and $g(\cdot)$ are differentiable with $s(g(0)) = s(\bar{p})$, the limit of (348) as $\alpha \to 0$ equals

$$\frac{\partial}{\partial \alpha}\big[w^i(s(g(\alpha)))\big]_{\alpha=0},$$

which, by the chain rule, equals

$$\nabla w^i(s(\bar{p})) \cdot Ds(\bar{p}) \nabla g(0). \tag{349}$$

The first and second of these partial derivatives is given directly by the functions $w^i(\cdot)$ and $s(\cdot)$. We will now find the third: $\nabla g(0) = (\nabla \hat{g}(0), 0)$.

Recall the construction of $\hat{g}(\alpha)$ in (339)–(341). By the IFT,

$$\nabla \hat{g}(0) = -\Big(D\underline{\hat{z}}_{\hat{p}}(\hat{\bar{p}}, 0)\Big)^{-1} D\underline{\hat{z}}_{\alpha}(\hat{\bar{p}}, 0), \tag{350}$$

where $\underline{\hat{z}}(\hat{p}, \alpha)$ is defined as in (340) and $D\underline{\hat{z}}_{\hat{p}}(\hat{\bar{p}}, 0)$ and $D\underline{\hat{z}}_{\alpha}(\hat{\bar{p}}, 0)$ are the Jacobians of $\underline{\hat{z}}(\cdot)$ with respect to $\hat{p}$ and α, respectively, evaluated at $(\hat{\bar{p}}, 0)$.

Element ℓ, k of $D\underline{\hat{z}}_{\hat{p}}(\hat{\bar{p}}, 0)$ (defined for $\ell, k < L$) equals

$$\frac{\partial \bar{x}_\ell(\check{\hat{\bar{p}}}, \hat{\bar{\pi}})}{\partial p_k} + \frac{\partial \bar{x}_\ell(\check{\hat{\bar{p}}}, \hat{\bar{\pi}})}{\partial \pi} \frac{\partial \underline{\hat{\pi}}(\hat{\bar{p}}, 0)}{\partial \hat{p}_k} - \frac{\partial s_\ell(\check{\hat{\bar{p}}})}{\partial p_k}. \tag{351}$$

Element ℓ $(< L)$ of $D\underline{\hat{z}}_{\alpha}(\hat{\bar{p}}, 0)$ equals

$$\frac{\partial \bar{x}_\ell(\check{\hat{\bar{p}}}, \hat{\bar{\pi}})}{\partial \pi} \frac{\partial \underline{\hat{\pi}}(\hat{\bar{p}}, 0)}{\partial \alpha} + dx^i_\ell, \tag{352}$$

where

$$dx^i \equiv x^i(\bar{p}, \bar{\pi}) - \bar{x}^i(\bar{p}, \bar{\pi}).$$

Likewise, recalling the construction of $\underline{\hat{\pi}}(\hat{p}, \alpha)$ following (339), and letting

$$\hat{\bar{\pi}} \equiv \pi(\check{\hat{\bar{p}}}, \bar{x}(\cdot)) \ \ \big(= \bar{\pi}/\bar{p}_L, \ = \underline{\hat{\pi}}(\hat{\bar{p}}, 0)\big),$$

the IFT gives us

$$\frac{\partial \underline{\hat{\pi}}(\hat{\bar{p}}, 0)}{\partial \hat{p}_k} = -\frac{1}{\bar{p} \cdot \hat{\bar{\delta}}} \sum_{\ell=1}^{L} \bar{p}_\ell \Big(\frac{\partial \bar{x}_\ell(\check{\hat{\bar{p}}}, \hat{\bar{\pi}})}{\partial p_k} - \frac{\partial s_\ell(\check{\hat{\bar{p}}})}{\partial p_k} \Big) \ \ (k < L), \tag{353}$$

$$\frac{\partial \underline{\hat{\pi}}(\hat{\bar{p}}, 0)}{\partial \alpha} = -\frac{1}{\bar{p} \cdot \hat{\bar{\delta}}} \ \bar{p} \cdot dx^i, \tag{354}$$

where

$$\bar{\delta} \equiv \nabla_\pi \bar{x}(\bar{p}, \bar{\pi}),$$
$$\hat{\bar{\delta}} \equiv \nabla_\pi \bar{x}(\check{\hat{\bar{p}}}, \hat{\bar{\pi}}) \ \ \big(= \bar{p}_L \bar{\delta}\big).$$

The denominator terms $\bar{p} \cdot \hat{\bar{\delta}}$ must be nonzero by the local regularity of $\bar{x}(\cdot)$.

Substituting (353) and (354) into (351) and (352), expression (350) can be rewritten

$$\nabla \hat{g}(0) = \Big[M\mathcal{I}\big(\hat{\bar{\delta}} \times \bar{p} - (\bar{p} \cdot \hat{\bar{\delta}}) I_L\big) \Big] dx^i,$$

and so

$$\nabla g(0) = \begin{bmatrix} M\mathcal{I}\big(\hat{\bar{\delta}} \times \bar{p} - (\bar{p} \cdot \hat{\bar{\delta}}) I_L\big) \\ \mathbb{0}_L^T \end{bmatrix} dx^i, \tag{355}$$

where

$$M \equiv \Big(\mathcal{I}\Big((\bar{p}\cdot\bar{\delta})\big(D\bar{x}_p(\check{\hat{\bar{p}}},\hat{\bar{\pi}}) - Ds(\check{\hat{\bar{p}}})\big) - D(\bar{\delta})\big(D_p\bar{x}(\check{\hat{\bar{p}}},\hat{\bar{\pi}}) - Ds(\check{\hat{\bar{p}}})\big)\bar{p}\times \mathbb{1}_L\Big)\mathcal{I}^T\Big)^{-1}.$$

Let G denote the matrix coefficient on dx^i in (355). Having established the linearity of $\nabla g(0)$ in dx^i, we will now provide an alternative characterization of G.

Define

$$\begin{aligned}
\underline{\pi}(p,\alpha) &\equiv \pi\big(p, \bar{x}(\cdot) + \alpha(x^i(\cdot) - \bar{x}^i(\cdot))\big) \ \ \big(= p_L\,\hat{\underline{\pi}}(\hat{p},\alpha)\big),\\
\underline{\bar{x}}(p,\alpha) &\equiv \bar{x}\big(p, \underline{\pi}(p,\alpha)\big),
\end{aligned}$$

and define $\underline{\bar{x}}^i(p,\alpha)$ and $\underline{x}^i(p,\alpha)$ likewise. Since $\hat{\underline{\pi}}(\hat{p},\alpha)$, as defined following (339), is differentiable and $\pi(p,x(\cdot))$ is h.o.d. 1 in p, $\underline{\pi}(\cdot)$ is differentiable. Thus $\underline{\bar{x}}(p,\alpha)$, $\underline{\bar{x}}^i(p,\alpha)$, and $\underline{x}^i(p,\alpha)$ are differentiable.

Now observe that, for all $\alpha \le 1/\underline{n}$, the following hold exactly:

$$\begin{aligned}
\underline{\bar{x}}(g(\alpha),\alpha) + \alpha(\underline{x}^i(g(\alpha),\alpha) - \underline{\bar{x}}^i(g(\alpha),\alpha)) &= s(g(\alpha)),\\
\underline{\bar{x}}(\bar{p},0) &= s(\bar{p})\\
\implies s(g(\alpha)) - s(\bar{p}) - \big(\underline{\bar{x}}(\bar{p},0) - \underline{\bar{x}}(g(\alpha),\alpha)\big) &= \alpha(\underline{x}^i(g(\alpha),\alpha) - \underline{\bar{x}}^i(g(\alpha),\alpha)). \quad (356)
\end{aligned}$$

Dividing both sides of (356) by α and taking the limit as $\alpha \to 0$, we have

$$-Dz(\bar{p})\nabla g(0) + \bar{\delta}\Big(\frac{\partial\underline{\pi}(\bar{p},0)}{\partial\alpha}\Big) = dx^i. \quad (357)$$

Because $\nabla g(0) = G dx^i$ and $\partial \underline{\pi}(\bar{p}, 0)/\partial \alpha = \partial \hat{\underline{\pi}}(\hat{\bar{p}}, 0)/\partial \alpha$, we then have

$$-Dz(\bar{p})G dx^i - \frac{p \cdot dx^i}{p \cdot \bar{\delta}} \bar{\delta} = dx^i. \tag{358}$$

Thus, for $dx^i \perp \bar{p}$, we have

$$-Dz(\bar{p})G dx^i = dx^i. \tag{359}$$

Observe that $Ds(\bar{p})\bar{p} = 0$. If prices all rise in proportion to their current levels, then the price vector has simply been rescaled, and because $s(p)$ is h.o.d. 0, supply levels will not change. $D\underline{x}(\bar{p})\bar{p} = 0$ likewise, by the definition of $\underline{x}$ and the fact that $\chi(p)$ is h.o.d. 0. So it follows from the definition (294) of $z(p)$ that

$$Dz(\bar{p})\bar{p} = 0. \tag{360}$$

Also, for any dp *not* proportional to $\bar{p}$, we have $d\hat{p} = \hat{p} - \hat{\bar{p}} \neq \mathbb{0}_{L-1}$. Because $\bar{p}$ is a regular WES, $D\hat{z}(\hat{\bar{p}})$ is of full rank, so $D\hat{z}(\hat{\bar{p}})d\hat{p} \neq \mathbb{0}_{L-1}$. It follows that any marginal price-change dp not proportional to $\bar{p}$ induces a change to excess demands, and thus that $Dz(\bar{p})dp \neq \mathbb{0}_L$. Thus

$$\text{Rank}(Dz(\bar{p})) = L - 1. \tag{361}$$

Returning to (359), we can now conclude that G is a generalized inverse of $-Dz(\bar{p})$. Furthermore, from (358), at $dx^i = -\bar{\delta}$ we have

$$Dz(\bar{p})G\bar{\delta} + \bar{\delta} = \bar{\delta}. \tag{362}$$

So $G\bar{\delta}$ either equals $\mathbb{0}_L$ or is proportional to $\bar{p}$. But we know that the bottom row of G consists of zeroes, so the last entry of $G\bar{\delta}$ equals 0. So

$$G\bar{\delta} = \mathbb{0}_L. \tag{363}$$

We will now construct the unique generalized inverse G of $Dz(\bar{p})$ whose bottom row consists of zeroes and for which $G\bar{\delta} = \mathbb{0}_L$.

Let UNV^T denote a singular value decomposition of $Dz(\bar{p})$ with $N_{\ell\ell} \neq 0$ for $\ell < L$ and $N_{LL} = 0$. Because G is a generalized inverse of $Dz(\bar{p})$, we must by (361) have

$$G = V \begin{bmatrix} N_1^{-1} & A \\ B & C \end{bmatrix} U^T, \tag{364}$$

where N_1 is the $(L-1) \times (L-1)$ principal submatrix of N and A, B, and C are $(L-1) \times 1$, $1 \times (L-1)$, and 1×1 respectively.

Since V is invertible, (363) reduces to

$$\begin{bmatrix} N_1^{-1} & A \\ B & C \end{bmatrix} \bar{\bar{\delta}} = \mathbb{0}_L \tag{365}$$

$$\implies A_\ell = -\frac{\bar{\bar{\delta}}_\ell}{\bar{\bar{\delta}}_L} \frac{1}{N_{\ell\ell}}, \quad \ell < L, \text{ and} \tag{366}$$

$$C = -\frac{1}{\bar{\bar{\delta}}_L} \sum_{\ell=1}^{L-1} B_\ell \bar{\bar{\delta}}_\ell, \tag{367}$$

where $\bar{\bar{\delta}} \equiv U^T \bar{\delta}$.

We will now impose the constraint that the bottom row of G consists of zeroes.

Since U^T is invertible, the bottom row of G consists of zeroes iff the bottom row of

$$V\begin{bmatrix} N_1^{-1} & A \\ B & C \end{bmatrix} \tag{368}$$

consists of zeroes. This in turn implies

$$B_\ell = -\frac{V_{L\ell}}{V_{LL}}\frac{1}{N_{\ell\ell}}, \quad \ell < L. \tag{369}$$

By (367), this also gives us C. G is thus fully constructed.

Substituting Gdx^i for $\nabla g(0)$ into (349), we have $\psi^i \cdot dx^i$, where

$$\psi^i = \big(Ds(\bar{p})G\big)^T \nabla w^i(s(\bar{p})). \tag{370}$$

Thus the limit of (345) as $n \to \infty$ equals

$$\big(v^i(x^i(\bar{p},\bar{\pi})) + \psi^i \cdot x^i(\bar{p},\bar{\pi})\big) - \big(v^i(\bar{x}^i(\bar{p},\bar{\pi})) + \psi^i \cdot \bar{x}^i(\bar{p},\bar{\pi})\big). \tag{371}$$

So $(\bar{p}, \{\bar{x}^i(\cdot)\})$ is a CESE iff $\bar{x}^i(\bar{p},\bar{\pi})$ maximizes $v^i(x^i(\bar{p},\bar{\pi})) + \psi^i \cdot x^i(\bar{p},\bar{\pi})$, among feasible $x^i(\bar{p},\bar{\pi})$, for all i.

B.4 Proof of Proposition 5

Continuous differentiability of demand given $\psi \not\ll 0$

Given $\psi^i \not\ll 0$ for all i, no individual is ever satiated. Thus $\pi(p,\psi) = p \cdot y(p)$ and may be denoted $\pi(p)$, and $\chi^i_{[\psi]}(\cdot)$ depends only on ψ^i and may be denoted $\chi^i_{[\psi^i]}(\cdot)$. We will now show that $\chi^i_{[\psi^i]}(p)$ is defined and locally $\mathcal{C}^1$ in p for all i, $\psi^i \not\ll 0$, $p \in P$.

By definition, $\chi^i_{[\psi^i]}(p)$ is the value of x^i that maximizes

$$\tilde{u}^i_{[\psi^i]}(x^i) = v^i(x^i) + \psi^i \cdot x^i \tag{372}$$

at the given p, subject to i's budget constraint. It follows from the strict concavity of $v^i(\cdot)$ that (372) is continuous and strictly quasiconcave in x^i. Since i's budget set is compact for each $p \gg 0$, and $p \gg 0$ for all $p \in P$, $\chi^i_{[\psi^i]}(p)$ exists and is unique.

Letting $\tilde{x}^i_{[\psi^i]}(\cdot)$ denote i's demand as a function of prices and wealth B^i, $\tilde{x}^i_{[\psi^i]}(p, B^i)$ is $\mathcal{C}^1$ with respect to both arguments around (p, B^i) if

$$H \equiv \begin{bmatrix} D^2\tilde{u}^i_{[\psi^i]}\big(\tilde{x}^i_{[\psi^i]}(p, B^i)\big) & p \\ p^T & 0 \end{bmatrix} \tag{373}$$

is nonsingular (Kreps (2012), Proposition 11.10). ($\psi^i \not\ll 0$ implies the required local non-satiation.) Given an $(L+1)$-vector d, the bottom row of H and $p \gg 0$ ensure that $[Hd]_{L+1} = 0$ only if the first L entries of d are all zero or contain both positives and negatives. By additive separability and since

$$\frac{\partial^2 \tilde{u}^i_{[\psi^i]}}{\partial (x^i_\ell)^2}\big(x^i\big) = v^{i\prime\prime}_\ell(x^i_\ell) < 0 \quad \forall x^i, \tag{374}$$

$[Hd]_\ell = 0$ for $\ell \leq L$ only if d_ℓ and d_{L+1} are of the same sign. So $[Hd] = 0$ iff $d = 0$.

Since

$$B^i = p \cdot e^i + \theta^i \pi(p),$$

$x^i_{[\psi^i]}(p, \pi)$ is $\mathcal{C}^1$ in p. It then follows from $\chi^i_{[\psi^i]}(p) \equiv x^i_{[\psi^i]}(p, \pi(p))$, the definition of $\pi(p)$, and the continuous differentiability of supply across P that $\chi^i_{[\psi^i]}(p)$ is $\mathcal{C}^1$ in p under the stated conditions.

Existence of WES given $\psi \not\ll 0$

Our assumptions guarantee that, unless $\psi^i \ll 0 \; \forall i$,

a) $s(\cdot)$ is uniformly smooth and quasi-monotonic on a compact, convex set of price vectors $P \subset \Delta^{L-1}$;

b) $v^i(\cdot)$ is additively separable, strictly increasing, differentiable, and weakly concave for all i;

c) $x_{[\psi]}(p, \pi) \gg 0 \; \forall p \gg 0$ (because for each good ℓ, $\exists i : \lim_{x^i_\ell \to 0} v^{i\prime}(x^i_\ell) = \infty, \; e^i \neq 0$);

d) $\exists i : \psi^i \not\ll 0, \; \theta^i > 0$; and

e) $x_{[\psi]}(\cdot)$ is defined, admissible, and continuous.

We will now show that there is a WES of $(\mathcal{E}, x_{[\psi]}(\cdot))$ and that all WES p satisfy $\tilde{p} \in P^\circ$. (The proof roughly follows the pattern of the proof of Mas-Colell et al. (1995), Proposition 17.C.1.)

Under the conditions above, for simplicity fix $x(\cdot) = x_{[\psi]}(\cdot)$, and drop the ψ subscripts on other terms likewise. By (d), $x(\cdot)$ is nonsatiated, so $\pi(p, x(\cdot))$ is defined for all p.

Consider the excess demand function

$$z(p) = x(p, \pi(p, x(\cdot))) - s(p).$$

We will first show that $z(p)$ is continuous in p throughout P. This will follow directly from showing that $\pi(p, x(\cdot))$ is continuous in p.

Recall from (290) that

$$\pi(p, x(\cdot)) = \pi : \ Z(p, \pi, x(\cdot)) = 0,$$

where

$$Z(p, \pi, x(\cdot)) \equiv p \cdot x(p, \pi) - p \cdot s(p).$$

Suppose by contradiction that $\pi(p, x(\cdot))$ is not continuous in p. Then there exists a $\bar{p}$ and a $\gamma > 0$ such that, for all $\epsilon > 0$,

$$\exists p \in \mathcal{N}_\epsilon(\bar{p}) \ : \ \pi(p, x(\cdot)) \notin \big(\pi(\bar{p}, x(\cdot)) - \gamma, \pi(\bar{p}, x(\cdot)) + \gamma\big).$$

For every natural $n > 0$, we can choose a $p^n \in \mathcal{N}_{1/n}(\bar{p})$ such that either

$$\pi(p^n, x(\cdot)) \geq \pi(\bar{p}, x(\cdot)) + \gamma \text{ or}$$
$$\pi(p^n, x(\cdot)) \leq \pi(\bar{p}, x(\cdot)) - \gamma.$$

There must thus be either an infinite subset ν of the naturals such that

$$\pi(p^n, x(\cdot)) \geq \pi(\bar{p}, x(\cdot)) + \gamma \ \ \forall n \in \nu \tag{375}$$

or one such that

$$\pi(p^n, x(\cdot)) \leq \pi(\bar{p}, x(\cdot)) - \gamma \ \ \forall n \in \nu. \tag{376}$$

Let ν be an infinite subset of the naturals such that (375) holds, and consider

the sequence $\{(p^n, \pi(\bar{p}, x(\cdot)) + \gamma)\}_{n\in\nu}$. Because $Z(p, \pi, x(\cdot))$ is strictly increasing in π, and $\pi : Z(p^n, \pi, x(\cdot)) = 0$ is no less than $\pi(\bar{p}, x(\cdot)) + \gamma$, we have

$$Z\big(p^n, \pi(\bar{p}, x(\cdot)) + \gamma, x(\cdot)\big) \leq 0 \ \ \forall n \in \nu. \tag{377}$$

Because $Z(p, \pi, x(\cdot))$ is continuous in p, however, and because $p^n \to \bar{p}$,

$$\{Z\big(p^n, \pi(\bar{p}, x(\cdot)) + \gamma, x(\cdot)\big)\}_{n\in\nu} \to Z(\bar{p}, \pi(\bar{p}, x(\cdot)) + \gamma) > 0, \tag{378}$$

with the inequality holding because $Z(\bar{p}, \pi(\bar{p}, x(\cdot)) + \gamma) = 0$ and $Z(p, \pi, x(\cdot))$ strictly increases in π. (377) contradicts (378), so there is no infinite subset of the naturals such that (375) holds.

Analogous reasoning proves that there is no infinite subset of the naturals such that (376) holds.

Therefore $\pi(p, x(\cdot))$ is continuous in p. So $z(p)$, as the composition of functions continuous throughout P, is also continuous in p throughout P.

We will now construct a correspondence $f(\cdot)$ from P to P, show that it has a fixed point $\bar{p}$, and then show that $\bar{p}$ is a WES.

For $p \in P^\circ$, let

$$f(p) = \{q \in P \ : \ \tilde{z}(p) \cdot q \geq \tilde{z}(p) \cdot q' \ \ \forall q' \in P\},$$

where

$$\tilde{z}(p) \equiv (I_L s(p))^{-1} z(p).$$

For $p \in P^c$, $\{1, ..., L\}$ can be partitioned into two nonempty sets of goods: $L_0(p) \equiv$

$\{\ell : s_\ell(p) = 0\}$ and its complement $L_1(p)$. Let

$$Q(p) \equiv \cup_{d \in \Delta^{L-1}_{L_0(p)}} \operatorname*{argmax}_{q \in P} q \cdot d.$$

That is, $Q(p)$ is the set of $q \in P$ that maximize a weighted average of the prices of the goods in $L_0(p)$. By the compactness of P, $Q(p) \neq \emptyset$. By the quasi-monotonicity of $s(\cdot)$, $p \notin Q(p)$.

We will now show that $Q(p)$ is closed. Given a sequence $q^n \to q^*$ with $q^n \in Q(p)\ \forall n$, since $\Delta^{L-1}_{L_0(p)}$ is compact, there is a convergent sequence $d^n \to d^*$ with $d^n \in \Delta^{L-1}_{L_0(p)}\ \forall n$, $d^* \in \Delta^{L-1}_{L_0(p)}$, and

$$q^n \in \operatorname*{argmax}_{q \in P} q \cdot d^n \ \ \forall n. \tag{379}$$

If $q^* \notin \operatorname{argmax}_{q \in P} q \cdot d^*$, then, since P is compact, there is a $q' \in P, \notin \Delta^{L-1}_{L_0(p)}$ with $q' \cdot d^* > q^* \cdot d^*$. Then by the continuity of $q \cdot d$ in q and d, we have $q' \cdot d^n > q^n \cdot d^n$ for n sufficiently large, contradicting (379).

Choose a homeomorphism $\phi(\cdot)$ between P^c and the unit $(L-1)$-sphere. Let

$$\epsilon(p) \equiv \min_{q \in Q(p)} ||\phi(q) - \phi(p)|| > 0,$$
$$h(p) \equiv \{p' \in P^c \ : \ ||\phi(p') - \phi(p)|| < \epsilon(p)\}.$$

Finally, let

$$f(p) = P^c \backslash h(p).$$

For all $p \in P$, $f(p)$ is a topological disc. If $p \in P^\circ$, this follows from the fact that $f(p)$ is a level set of a linear function defined on the compact, convex set P, and so

is compact and convex. If $p \in P^c$, this follows from the facts that the boundary of a closed disc (as P is) is a topological sphere, that $h(p)$ is a hole in P^c (since the image $\phi(h(p))$ is a hole in the unit $(L-1)$-sphere), and that a topological sphere with a hole removed is a topological disc.

To show that $f(\cdot)$ is upper hemicontinuous, consider a sequence $(p^n, q^n) \to (p, q)$, all within P, with $q^n \in f(p^n)\ \forall n$.

If $p \in P^\circ$, then $p^n \in P^\circ$ for n sufficiently large. It follows from $q^n \cdot \tilde{z}(p^n) \geq q' \cdot \tilde{z}(p^n)\ \forall q' \in P$ and the continuity of $\tilde{z}(\cdot)$ across P° that $q \in f(p)$.

If $p \in P^c$, then there is either a subsequence whose p-elements are contained in P°, or one whose p-elements are contained in P^c. Let such a subsequence be indexed by m.

In the first case, by the continuity of $z(\cdot)$ and $s(\cdot)$ on P, and the fact that $\chi(p) \gg 0\ \forall p \in P$, we have $\lim_{m\to\infty} \tilde{z}_\ell(p^m) = \infty$ if $\ell \in L_0(p)$, with the limit finite otherwise. If $q \notin Q(p)$, there is a $q' \in P$ with $q'_\ell \geq q_\ell$ for all $\ell \in L_0(p)$ and $q'_\ell > q_\ell$ for some $\ell \in L_0(p)$. Thus for m sufficiently large, $\tilde{z}(p^m) \cdot q' > \tilde{z}(p^m) \cdot q^m$, contradicting the stipulation that $q^m \in f(p^m)$. So $q \in Q(p) \subset f(p)$.

In the second case, we can choose a subsequence $\{(p^m, q^m)\}$ such that, for m sufficiently large, $L_0(p^m)$ is constant and, by the continuity of $s(\cdot)$ on P, a weak subset of $L_0(p)$. Then for m sufficiently large, $q^m \in Q(p^m) \subseteq Q(p) \subseteq f(p)$. Since $Q(p)$ is closed, $q \in Q(p) \subset f(p)$.

We have established that $f(\cdot)$ is a disc-valued, upper hemicontinuous correspondence from a nonempty, compact, convex set P to itself. By Kakutani's fixed point theorem, there is a $\bar{p} \in P$ with $\bar{p} \in f(\bar{p})$.

We cannot have $\bar{p} \in P^c$, since $s(\bar{p}) \not\gg 0$ but $\chi(\bar{p}) \gg 0$ for $\bar{p} \in P^c$. Given $\bar{p} \in P^\circ$,

we must have $z(\bar{p}) = 0$. Otherwise there are ℓ, k with $z_\ell(\bar{p}) > 0$, $z_k(\bar{p}) < 0$. But $z(\bar{p}) \cdot \bar{p} = 0$, by construction of the profit function; and since $\bar{p} \in P^\circ$ there is an $\epsilon > 0$ such that $p \equiv \bar{p} + \epsilon(a_\ell - a_k) \in P$ and $\tilde{z}(\bar{p}) \cdot p > 0$, where in general a_ℓ denotes the unit vector with a one in place ℓ. This contradicts $\bar{p} \in f(\bar{p})$. So $\bar{p}$ is a WES.

Also, more generally, observe that all WES must be in P°, to ensure positive supply.

Regularity of all WES given $\psi \not\ll 0$

Let $\lambda^i_{[\psi^i]}(p)$ denote the marginal utility of nominal wealth for i at p given ψ^i. Since for each ℓ

$$\lambda^i_{[\psi^i]}(p) = \frac{1}{p_\ell}\Big(v^{i\prime}_\ell(\chi^i_{[\psi^i]\ell}(p)) + \psi^i_\ell\Big), \tag{380}$$

the continuous differentiability of $\chi^i_{[\psi^i]}(\cdot)$ and twice continuous differentiability of $v^i_\ell(\cdot)$ imply that $\lambda^i_{[\psi^i]}(\cdot)$ is $\mathcal{C}^1$ in p under the stated conditions as well.

Let

$$\tilde{\Psi} \equiv \big\{\psi \,:\, \psi^i \not\ll 0 \;\forall i,\; \psi^i = \psi^j \;\forall i, j\big\}$$

denote the set of ψ with ψ^i identical and not entirely negative for all i. We will now show that $Dz_{[\psi]}(p)dp \neq 0$ for all $p \in P$, $dp \not\propto p$, and $\psi \in \tilde{\Psi}$. This will establish that $D\hat{z}_{[\psi]}(\hat{p})$ is invertible for all (p, ψ) with $\tilde{p} \in P$ and $\psi \in \tilde{\Psi}$.

The lower Inada condition rules out corner solutions, so $\chi^i_{[\psi^i]}(p)$ satisfies, for the given p and ψ^i,

$$Dv^i\big(\chi^i_{[\psi^i]}(p)\big)^T + \psi^i = \lambda^i_{[\psi^i]}(p)p. \tag{381}$$

Differentiating (381) with respect to p yields

$$D^2v^i\big(\chi^i_{[\psi^i]}(p)\big)D\chi^i_{[\psi^i]}(p) = pD\lambda^i_{[\psi^i]}(p) + \lambda^i_{[\psi^i]}(p)I_L$$
$$\implies D\chi^i_{[\psi^i]}(p) = D^2v^i\big(\chi^i_{[\psi^i]}(p)\big)^{-1}\Big(pD\lambda^i_{[\psi^i]}(p) + \lambda^i_{[\psi^i]}(p)I_L\Big). \qquad (382)$$

Fixing p, choose a price-change vector $dp \not\propto p$, and let

$$\tilde{dp}_{[\psi^i]} \equiv dp - \frac{D\lambda^i_{[\psi^i]}(p)\cdot dp}{D\lambda^i_{[\psi^i]}(p)\cdot p}p.$$

Note that the denominator represents the change in the marginal utility of wealth for i per unit change of prices in the direction of p: that is, per unit of positive rescaling of the price vector. A positive rescaling of the price vector leaves χ^i unchanged, proportionally increasing both i's budget and all prices, proportionally lowering the marginal utility of nominal wealth (which positive by $\psi^i \gg 0$). So the denominator is negative, and in particular nonzero.

Since $D\chi^i_{[\psi^i]}(p)p = 0$, $D\chi^i_{[\psi^i]}(p)\tilde{dp}_{[\psi^i]} = D\chi^i_{[\psi^i]}(p)dp$. Also, by construction, $\tilde{dp}_{[\psi^i]}$ is an eigenvector of the second factor on the right-hand side of (382), with

$$\Big(pD\lambda^i_{[\psi^i]}(p) + \lambda^i_{[\psi^i]}(p)I_L\Big)\tilde{dp}_{[\psi^i]} = \lambda^i_{[\psi^i]}(p)\tilde{dp}_{[\psi^i]}.$$

Then, since the corresponding eigenvalue is $\lambda^i_{[\psi^i]}(p) > 0$, and since by additive separability and strict concavity $D^2v^i\big(\chi^i_{[\psi^i]}(p)\big)^{-1}$ is a diagonal matrix with negative entries, (382) gives us that $D\chi^i_{[\psi^i]}(p)dp$ is a vector whose entries have signs opposite to those of $\tilde{dp}_{[\psi^i]}$.

Since the normalized Jacobian of supply $D\hat{s}(\hat{p})$ is positive semidefinite, and since by the nonsingularity assumption $Ds(p)\tilde{dp}_{[\psi^i]} \neq 0$ for $\tilde{dp}_{[\psi^i]} \not\propto p$, $Ds(p)\tilde{dp}_{[\psi^i]}$ preserves

the sign of at least one nonzero entry of $\tilde{dp}_{[\psi^i]}$.

Given symmetry of preferences and budgets, $\psi \in \tilde{\Psi}$ implies that $D\chi_{[\psi]}(p)$ is proportional to $D\chi^i_{[\psi^i]}(p)$ for each i. This completes the proof that $Dz_{[\psi]}(p)dp \neq 0$ under the stated conditions.

Bounding ψ

Let $\tilde{dp}_{[\psi]}$ denote $\tilde{dp}_{[\psi^i]}$ for any i (given $\psi \in \tilde{\Psi}$ and thus ψ^i the same for all i), and let

$$\mathcal{P} \equiv \{dp : ||dp|| = 1, dp_L = 0\}$$

denote the set of price-change vectors that leave the price of good L fixed and whose absolute value is 1. We will now show that

$$\underline{m} \equiv \inf_{p \in P, dp \in \mathcal{P}, \psi \in \tilde{\Psi}} \left[\max_{\ell} \left| [-Dz_{[\psi]}(p)dp]_\ell \right| \right] > 0.$$

If not, there is a sequence $p^{(n)}, dp^{(n)}, \psi^{(n)}$ in $P \times \mathcal{P} \times \tilde{\Psi}$ such that

$$\lim_{n \to \infty} -Dz_{[\psi^{(n)}]}(p^{(n)})dp^{(n)} = 0. \tag{383}$$

Because P and $\mathcal{P}$ are compact, there is a subsequence of this sequence for which $p^{(n)}$ and $dp^{(n)}$ converge to some p^*, dp^* and limit (383) is maintained. Also, for each n in this subsequence, there is a unique $\kappa^{(n)} \in \mathbb{R}$ such that

$$\tilde{dp}^{(n)}_{[\psi^{(n)}]} = dp^{(n)} + \kappa^{(n)} p^{(n)}.$$

We can therefore choose a sequence—a subsequence of this subsequence of the original sequence—for which $p^{(n)}$ and $dp^{(n)}$ converge, limit (383) is maintained, and either

a) $\kappa^{(n)} \to \kappa^* \in (-\infty, \infty)$ or

b) $\kappa^{(n)} \to \infty$ or $\kappa^{(n)} \to -\infty$.

Choose such a sequence. Because $Ds(p)dp$ is continuous in both arguments, (383) implies

$$\lim_{n\to\infty} D\chi_{[\psi^{(n)}]}(p^{(n)})dp^{(n)} = \lim_{n\to\infty} Ds(p^{(n)})dp^{(n)} = Ds(p^*)dp^*.$$

So for each nonzero element of $Ds(p^*)dp^*$ (of which there is at least one, since $dp^* \not\propto p^*$), the sign of the corresponding element of $D\chi_{[\psi^{(n)}]}(p^{(n)})dp^{(n)}$ is the same for all sufficiently large n. But for each n, there is an ℓ such that $[Ds(p^{(n)})dp^{(n)}]_\ell$ is nonzero and of the same sign as $[\tilde{dp}^{(n)}_{[\psi^{(n)}]}]_\ell$, and so (equivalently) of opposite sign to $[D\chi_{[\psi^{(n)}]}(p^{(n)})dp^{(n)}]_\ell$. Let $\ell^{(n)} > 0$ denote the absolute value of the element(s) among these that is (are) largest in absolute value. (383) implies that $\ell^{(n)} \to 0$. We will now show that this is incompatible with both cases (a) and (b).

In case (a), $\ell^{(n)} \to 0$ implies

$$\lim_{n\to\infty} \tilde{dp}^{(n)}_{[\psi^{(n)}]} \cdot Ds(p^{(n)})\tilde{dp}^{(n)}_{[\psi^{(n)}]} = (dp^* + \kappa^* p^*)Ds(p^*)(dp^* + \kappa^* p^*) \leq 0.$$

This implies that $D\hat{s}(\hat{p})$ is not positive definite. This is impossible, since as the Jacobian of a supply function it is positive semidefinite and symmetric, and by assumption it is nonsingular.

In case (b), suppose $\kappa^{(n)} \to \infty$. Then there is an $\underline{n}$ such that $\tilde{dp}^{(n)}_{[\psi^{(n)}]} \gg 0 \; \forall n \geq \underline{n}$. So $D\chi_{[\psi^{(n)}]}(p^{(n)})dp^{(n)} \ll 0 \; \forall n \geq \underline{n}$. So each entry of $Ds(p^*)dp^*$ is weakly negative, with at least one entry strictly negative. This is impossible, since $p^* \cdot Ds(p^*)dp^* = 0$ (recalling that $p^* \cdot Ds(p^*) = 0$, by symmetry of $Ds(p) \; \forall p$). The $\kappa^{(n)} \to -\infty$ case is ruled out analogously.

Let a_ℓ denote the unit vector with a one in place ℓ. Suppose $dx^i = a_\ell$, and observe that $||dx^i|| = 1$. For any $p \in P$ and $\psi \in \tilde{\Psi}$, there is a unique dp with $dp_L = 0$ satisfying

$$-Dz_{[\psi]}(p)dp = dx^i$$

(where $G_{[\psi]}(p)$ is the generalized inverse of $-Dz_{[\psi]}(p)$ defined by (306), and $G_{[\psi]}(p)dx^i = dp + \kappa p$ for some κ). We must have $||dp|| \leq 1/\underline{m}$; if greater, then there is an entry of $-Dz_{[\psi]}(p)dp$ with absolute value greater than $\underline{m}(1/\underline{m}) = 1$, so

$$|| - Dz_{[\psi]}(p)dp|| = ||dx^i|| > 1,$$

a contradiction. Since

$$\begin{aligned} \psi^i_\ell(p,\psi)^T &= Dw^i(s(p))Ds(p)G(p,\psi)a_\ell \\ &= Dw^i(s(p))Ds(p)dp, \end{aligned} \tag{384}$$

where the second expression is continuous in p and dp, and since P and $\{dp : dp_L = 0, ||dp|| \leq 1/\underline{m}\}$ are compact, there is a bound $\bar{\bar{\psi}} \gg 0$ such that

$$\psi^i(p,\psi) \in [-\bar{\bar{\psi}}, \bar{\bar{\psi}}]^L \ \ \forall p \in P,\ \psi \in \tilde{\Psi}.$$

Furthermore, the additive separability of all utility functions guarantees that $\delta(p,\psi) \geq 0$, and $\psi \in \tilde{\Psi}$ implies condition (d) listed under "Existence of WES", which guarantees that $\delta(p,\psi) \neq 0$. There is thus for each i some $dx^i = \delta(p,\psi) > 0$ such that $G(p,\psi)x^i = 0$, and therefore such that $\psi^i(p,\psi) \cdot dx^i = 0$. We cannot have $\psi^i(p,\psi) \ll 0$ for any i. So, letting Ψ denote the subset of $\tilde{\Psi}$ with $\psi^i \in [-\bar{\bar{\psi}}, \bar{\bar{\psi}}]^L\ \forall i$,

we have

$$\psi(p,\psi) \in \Psi \ \ \forall p \in P,\, \psi \in \Psi. \tag{385}$$

Observe that Ψ is compact and a topological disc.

Existence of RCESE

At any price vector p, all consumers have the same budgets and quasi-utility functions, and thus buy the same baskets. Under this constraint, the strict quasiconcavity of $\tilde{u}^i_{\psi^i}(\cdot)$ for any ψ and the strict convexity of the production set imply that a unique aggregate supply vector s^* is efficient. By the first fundamental theorem of welfare economics, if $\bar{p}$ is a Walrasian equilibrium, $s(\bar{p}) = s^*$.

Suppose there are price vectors $p \in P^\circ$, $p' \neq p, \in P^\circ$ with $s(p) = s(p') = s^*$. Then for all $\alpha \in [0,1]$,

$$\underset{s}{\operatorname{argmax}}\ ((1-\alpha)p + \alpha p') \cdot s = s^*,$$

since s^* maximizes both $p \cdot s$ and $p' \cdot s$. Thus $Ds(p)(p' - p) = 0$, implying that $D\hat{s}(\hat{p})$ is singular at p. But since p and p' are both in P, we cannot have $p' - p \propto p$. By definition of uniform smoothness, it follows that there is a unique $p \in P^\circ$ with $s(p) = s^*$.

So there is a unique WES $\bar{p} \in P^\circ$ for each externality matrix $\psi \in \Psi$. Denote it by $p(\psi)$. We will now show that $p(\cdot)$ is $\mathcal{C}^1$ across Ψ.

First we will show that $\chi_{[\psi]}(p)$ is $\mathcal{C}^1$ in both arguments throughout $P \times \Psi$. Recall that we have already established that $\chi_{[\psi]}(p)$ is $\mathcal{C}^1$ in p for any $\psi \in \Psi$. Also, recall

that $\psi \in \Psi$ guarantees budget-exhaustion for all individuals, so that we can write $\chi^i_{[\psi^i]}(p)$, noting that $\chi^i_{[\cdot]}(\cdot)$ is continuous in prices, because $x^i_{[\psi^i]}(p,\pi)$ is continuous in (p,π) and $\pi = p \cdot y(p)$ is continuous in p.

Suppose that, for some i, $\chi^i_{[\psi^i]}(p)$ is not even continuous in (p,ψ) throughout $P \times \Psi$. Then there is a sequence $(p^{(n)}, \psi^{i(n)}) \to (p, \psi^i)$, all within $P \times \Psi$, such that $\chi^i_{[\psi^{i(n)}]}(p^{(n)})$ converges to some χ^{i*} (by the Bolzano-Weierstrass theorem and the fact that maximal demands for i are bounded across all budgets supported by $p \in P$) and

$$\chi^i_{[\psi^i]}(p) \neq \chi^{i*}. \tag{386}$$

Since utility-maximizing demands are single-valued, $\tilde{u}^i_{[\cdot]}(\cdot)$ is continuous in both arguments, and $\chi^i_{[\cdot]}(\cdot)$ is continuous in prices, (386) implies that, for n sufficiently large,

$$\tilde{u}^i_{[\psi^{i(n)}]}\big(\chi_{[\psi^i]}(p^{(n)})\big) > \tilde{u}^i_{[\psi^{i(n)}]}\big(\chi_{[\psi^{i(n)}]}(p^{(n)})\big),$$

contradicting the requirement that $\chi_{[\psi^{i(n)}]}(p^{(n)})$ be maximize $\tilde{u}^i_{[\psi^{i(n)}]}$ at prices $p^{(n)}$. So $\chi^i_{[\cdot]}(\cdot)$ is continuous in both arguments for all i, and the sum $\chi_{[\cdot]}(\cdot)$ is as well.

Given a marginal shift $d\psi^i$ to i's externality vector, holding prices and profits fixed, i's demands x^i maintain the two optimality conditions

1. $\frac{1}{p_\ell}\frac{\partial \tilde{u}^i}{\partial x^i_\ell}$ equal for all ℓ,
2. $p \cdot x^i = B^i$

if they shift marginally by dx^i satisfying

$$\begin{aligned} D^2v^i(x^i)dx^i + d\psi^i &= kp \\ \implies dx^i &= \left(D^2v^i(x^i)\right)^{-1}(kp - d\psi^i) \quad \text{for some } k, \text{ and} \qquad (387) \\ p \cdot dx^i &= 0. \qquad (388) \end{aligned}$$

Combining (387) and (388) yields

$$\begin{aligned} p \cdot \left(D^2v^i(x^i)\right)^{-1}(kp - d\psi^i) &= 0 \\ \implies k &= \frac{p \cdot \left(D^2v^i(x^i)\right)^{-1} d\psi^i}{p \cdot \left(D^2v^i(x^i)\right)^{-1} p}. \qquad (389) \end{aligned}$$

Substituting (389) into (387) and simplifying yields

$$dx^i = \left(1 - \frac{1}{p \cdot \left(D^2v^i(x^i)\right)^{-1} p}\right)\left(D^2v^i(x^i)\right)^{-1} d\psi^i. \qquad (390)$$

Since (390) is defined for all $p \gg 0$ (so for all $p \in P$) and x^i with $x^i_{[\psi^i]}(p, \pi)$ (for some $\psi \in \Psi$ and $\pi \geq 0$), $x^i_{[\psi^i]}(p, \pi)$ is differentiable in ψ^i on $P \times \Psi$, with the Jacobian given by the coefficient on $d\psi^i$ on the right-hand side of (390). Because $x^i_{[\psi^i]}(p, \pi) = \chi^i_{[\psi^i]}(p)$ is continuous in p and ψ^i, and $v^i(\cdot)$ is $\mathcal{C}^2$ in x^i, this Jacobian is continuous in (p, π). This establishes that the partial derivatives of $\chi^i_{[\psi^i]}(p)$ with respect to ψ^i exist and are continuous in (p, ψ^i).

The last step is to establish that the partial derivatives of $\chi^i_{[\psi^i]}(p)$ with respect to p are also continuous in (p, ψ^i) (and not just p in isolation). This follows from immediately from (382) and the continuous differentiability of $\lambda^i_{[\psi^i]}(p)$ in both arguments, which in turn follows from (380) and the continuity of $\chi^i_{[\psi]}(p)$ in both arguments.

So $\chi_{[\psi]}(p)$ is $\mathcal{C}^1$ in both arguments across $P \times \Psi$.

Because $s(\cdot)$ is $\mathcal{C}^1$ across P and $\chi_{[\cdot]}(\cdot)$ is $\mathcal{C}^1$ in both arguments across $P \times \Psi$, $z_{[\cdot]}(\cdot)$ is $\mathcal{C}^1$ in both arguments across $P \times \Psi$. So therefore is $\hat{z}_{[\cdot]}(\cdot)$, where, following (298),

$$\hat{z}_{[\psi]}(\hat{p}) \equiv \mathcal{I} z_{[\psi]}(\check{\hat{p}}). \tag{391}$$

Recall also that $D_{\hat{p}}\hat{z}_{[\psi]}(\hat{p})$ is invertible for any $(\hat{p}, \psi)$ for which $\hat{p}$ is a regular normalized WES given ψ. By the IFT, this establishes that if $p(\psi)$ is defined, $p(\cdot)$ is defined and continuous (in fact $\mathcal{C}^1$) throughout a neighborhood of ψ in Ψ. The continuity of $z_{[\cdot]}(\cdot)$ in both arguments then implies that $p(\psi)$ is defined and continuous for all $\psi \in \Psi$.

By (385) and the fact that $\tilde{p} \in P$ for all WES p,

$$\psi^*(\psi) \equiv \psi(p(\psi), \psi)$$

is a function from the nonempty compact topological disc Ψ to itself. We will now show that $\psi^*(\cdot)$ is continuous.

We will first show that $G(p(\psi), \psi)$ is continuous in ψ throughout Ψ.

Suppose it is not. Then there exist $\psi, \psi' \in \Psi$ such that

$$\lim_{\epsilon \to 0} G\big(p(\psi^{(\epsilon)}), \psi^{(\epsilon)}\big) \neq G\big(p(\psi), \psi\big),$$

where

$$\psi^{(\epsilon)} \equiv \psi + \epsilon(\psi' - \psi).$$

There must then be some $dx^i \in \mathbb{R}^L$ such that

$$\lim_{\epsilon \to 0} G\big(p(\psi^{(\epsilon)}), \psi^{(\epsilon)}\big) dx^i \neq G\big(p(\psi), \psi\big) dx^i. \tag{392}$$

Denote the left- and right-hand sides by g and g^* respectively. By construction, $g_L = g^*_L = 0$. There is therefore no $\kappa \neq 0$ such $g = g^* + \kappa p(\psi)$. Then since $z_{[.]}(\cdot)$ is locally $\mathcal{C}^1$ in both arguments around $(p(\psi), \psi)$, and since $p(\psi)$ is an RWES compatible with externality matrix ψ, (392) in turn implies

$$\lim_{\epsilon \to 0} -Dz_{[\psi^{(\epsilon)}]}\big(p(\psi^{(\epsilon)})\big) G\big(p(\psi^{(\epsilon)}), \psi^{(\epsilon)}\big) dx^i \neq \lim_{\epsilon \to 0} -Dz_{[\psi^{(\epsilon)}]}\big(p(\psi^{(\epsilon)})\big) G\big(p(\psi), \psi\big) dx^i.$$

This is impossible: the term in the limit on the left-hand side equals dx^i for all ϵ, by definition of G, and the right-hand limit equals dx^i because $p(\cdot)$ is continuous and $-z_{[.]}(\cdot)$ is locally $\mathcal{C}^1$ in both arguments around $(p(\psi), \psi)$.

By the definition of ψ, the continuous differentiability of $w^i(s)$ in s for all i, and the continuous differentiability of $s(p)$ in p throughout P, $\psi(\psi)$ is continuous. The function $\psi(\psi)$ is thus a continuous function from the nonempty compact topological disc Ψ to itself. By Brouwer's fixed point theorem, it has a fixed point. By construction, for any such fixed point $\bar{\psi}$, $(p(\bar{\psi}), \bar{\psi})$ is an RCESE.

B.5 Proof of Proposition 6

The proof largely mirrors, and refers to, that in Appendix B.4. Let

$$
\begin{aligned}
v_\ell^{i*} &\equiv \lim_{x_\ell^i \to \infty} v_\ell^{i\prime}(x_\ell^i), \\
\underline{v}_L^i(\psi^i) &\equiv \max\Big(0,\, \max_\ell \left(\psi_\ell^i + v_\ell^{i*}\right)\Big), \\
\tilde{\Psi} &\equiv \{\psi \,:\, \psi_L^i = 0\ \forall i,\ \ w^i(\cdot) = 0 \implies \psi^i = 0\ \forall i\}.
\end{aligned}
$$

Continuous differentiability of demand given $\psi \in \tilde{\Psi}$ and high $\{v_L^i\}, \{B^i\}$

If $\psi \in \tilde{\Psi}$ and $v_L^i > 0$, no individual is ever satiated. Thus $\pi(p, \psi) = p \cdot y(p)$ and may be denoted $\pi(p)$, and $\chi_{[\psi]}^i(\cdot)$ depends only on ψ^i and may be denoted $\chi_{[\psi^i]}^i(\cdot)$.

We will now show that for all (p, ψ) with $p \in P$ and $\psi \in \tilde{\Psi}$, for each i, if $v_L^i > \underline{v}_L^i(\psi^i)$, there is a budget $\underline{B}^i(p, \psi^i, v_L^i)$ such that if $B^i > \underline{B}^i(p, \psi^i, v_L^i)$, then $\chi_{[\psi^i]}^i(p)$ is defined and locally $\mathcal{C}^1$ in p.

Recall that $\chi_{[\psi^i]}^i(p)$ is the value of x^i that maximizes $\tilde{u}_{[\psi^i]}^i$ (see (372)) at the given p, subject to i's budget constraint. It follows from the quasilinearity of $v^i(\cdot)$ in L, the strict concavity of $v_\ell^i(\cdot)\ \forall \ell < L$, and $\psi_L^i = 0$ that (372) is continuous and strictly quasiconcave in x^i and thus that $\chi_{[\psi^i]}^i(p)$ is defined.

Define $\tilde{x}_{[\psi^i]}^i(\cdot)$ as in Appendix B.4. The lower Inada condition on $v_\ell^i(\cdot)$ for $\ell < L$ ensures that $\tilde{x}_{\ell[\psi^i]}^i(\cdot)(p, B^i) > 0$ for all $p \gg 0$, $B^i > 0$, $\ell < L$. The $v_L^i > \underline{v}_L^i(\psi)$ condition ensures that for each i and each $p^* \gg 0$ there is a budget

$$\underline{B}^i(p, \psi^i, v_L^i)$$

such that if $B^{i*} > \underline{B}^i(p, \psi^i, v_L^i)$, $\tilde{x}_{L[\psi^i]}^i(p, B^i) > 0$ for all (p, B^i) near (p^*, B^{i*}). The $v_L^i > 0$ and $\psi_L^i = 0$ conditions imply local non-satiation.

Then $\tilde{x}^i_{[\psi^i]}(p, B^i)$ is $\mathcal{C}^1$ with respect to both arguments around (p, B^i) if (373) is nonsingular. Given an $(L+1)$-vector d, the bottom row of H and $p \gg 0$ ensure that $[Hd]_{L+1} = 0$ only if the first L entries of d are all zero or contain both positives and negatives. By additive separability and since (374) holds for all $\ell < L$, $[Hd]_\ell = 0$ for $\ell < L$ only if d_ℓ and d_{L+1} are of the same sign. Also, since

$$\frac{\partial^2 \tilde{u}^i_{[\psi^i]}}{\partial (x^i_L)^2}\big(x^i\big) = 0 \quad \forall x^i,$$

$[Hd]_L = 0$ iff $d_L = 0$. So $[Hd] = 0$ iff $d = 0$.

It then follows as in Appendix B.4 that $\chi^i_{[\psi^i]}(p)$ is locally $\mathcal{C}^1$ in p if $\psi \in \tilde{\Psi}$, $v^i_L > \underline{v}^i_L(\psi^i)\ \forall i$, and $B^i > \underline{B}^i(p, \psi^i, v^i_L)\ \forall i$.

Finally, fixing v^i_L, let us show that $\underline{B}^i(\cdot)$ is continuous in its first two arguments, under the constraint that $\underline{v}^i_L(\psi) < v^i_L$. $\underline{B}^i(p, \psi, v^i_L)$ is the minimum value of B^i that sets

$$v^{i\prime}_1(\tilde{x}^i_{1[\psi^i]}(p, B^i)) + \psi^i_1 = v^i_L, \tag{393}$$

and so the unique value of B^i that would maintain (393) if good L could not be purchased. Dropping good L from consideration, $\tilde{x}^i_{[\psi^i]}(\cdot)$ is $\mathcal{C}^1$ in prices and wealth, by the same proof of continuous differentiability as in Appendix B.4 (now with $L-1$ goods). A proof precisely analogous to that establishing that $\chi^i_{[\cdot]}(\cdot)$ is $\mathcal{C}^1$ in both arguments (see (386)–(390) and the surrounding discussion) then establishes that $\tilde{x}^i_{[\cdot]}(\cdot)$ is $\mathcal{C}^1$ in all three arguments (again, under the restriction that only the first $L-1$ goods may be purchased). By the IFT, $\underline{B}^i(\cdot)$ is continuous in p and ψ^i.

By continuity in p, we can define

$$\underline{b}^i(\psi, v_L^i) \equiv \min_{p \in P} \underline{B}^i(p, \psi, v_L^i).$$

Existence of WES given $\psi \in \tilde{\Psi}$ and high $\{v_L^i\}, \{B^i\}$

Our assumptions guarantee that, if $\psi_L^i = 0 \ \forall i$, $v_L^i > \underline{v}_L^i(\psi)$, and $B^i > \underline{b}^i(\psi, v_L^i)$, conditions (a)–(e) under the "Existence of WES given $\psi \not\ll 0$" heading of Appendix B.4 are satisfied.

The proof that $z_{[\psi]}(\cdot)$ is defined and continuous throughout P is as in Appendix B.4. If $s(\cdot)$ is quasi-monotonic on P, the proof that a WES exists is also as in Appendix B.4. Suppose therefore that for all ℓ,

$$\exists i : w^i(\cdot) = 0, e_\ell^i > 0 \tag{394}$$

(and do not assume that $s(\cdot)$ is quasi-monotonic on P).

Because the production possibility set (and thus the supply possibility set) is compact,

$$\exists \bar{s} > 0 : s_\ell(p) < \bar{s} \ \ \forall \ell \ \ \forall p \gg 0. \tag{395}$$

It then follows from non-negativity of demands that

$$z_\ell(p) > -\bar{s} \ \ \forall \ell \ \ \forall p \gg 0. \tag{396}$$

Consider a sequence of positive prices $\{p^n\} \to p \neq 0$ such that $p_k = 0$ for some k. Choose an ℓ with $p_\ell > 0$, and choose an individual i such that condition (394) is

satisfied for ℓ, so that

$$\lim_{n\to\infty} p^n \cdot e^i + \theta^{i\ell}\pi(p^n) > 0.$$

It follows from the additive separability and strict monotonicity of $v^i(\cdot)$ that

$$\lim_{n\to\infty} \chi^i_{k[\psi]}(p^n) = \infty.$$

Since demands by other individuals cannot be negative, and supply levels are bounded above, we have

$$\lim_{n\to\infty} \left\{ \max_k \left(z_{k[\psi]}(p^n) \right) \right\} = \infty. \tag{397}$$

It follows by Mas-Colell et al. (1995), Proposition 17.C.1 that $\exists \bar{p} \gg 0 : z(\bar{p}) = 0$. By definition, $\bar{p}$ is a WES.

Since demands are positive, $\tilde{\bar{p}} \in P^\circ$.

Regularity of all WES given $\psi \in \tilde{\Psi}$ and high $\{v^i_L\}, \{B^i\}$

If $\psi^i_L = 0\ \forall i$, $v^i_L > \underline{v}^i_L(\psi)$, and $B^i > \underline{b}^i(\psi, v^i_L)$, then $p \in P$ implies $\chi^i_{L[\psi^i]}(p) > 0\ \forall i$. Thus the marginal utility of nominal wealth for i is

$$\frac{v^{i\prime}_\ell(\chi^i_{[\psi]}(p))}{p_\ell} = \frac{v^i_L}{p_L} \quad \forall \ell. \tag{398}$$

For $\ell < L$, differentiating with respect to p_ℓ yields

$$\frac{\partial \chi^i_{\ell[\psi^i]}(p)}{\partial p_\ell} = \frac{v^i_L}{p_L\, v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} < 0. \tag{399}$$

For $k < L$, $k \neq \ell$, differentiating with respect to p_k yields

$$\frac{\partial \chi^i_{\ell[\psi^i]}(p)}{\partial p_\ell} = 0. \tag{400}$$

The upper-left $(L-1) \times (L-1)$ submatrix of $D\chi^i_{[\psi^i]}(p)$ thus has negative entries on the diagonal and zeroes elsewhere.

For completeness, to verify that $D\chi^i_{[\psi^i]}(p)$ is defined, we can identify its bottom row and right column as follows. For $\ell < L$, differentiating the budget exhaustion condition

$$p \cdot \chi^i_{[\psi^i]}(p) = p \cdot e^i + \theta^i (p \cdot y(p))$$

with respect to p_ℓ yields

$$\sum_{k=1}^{L} p_k \frac{\partial \chi^i_{k[\psi^i]}(p)}{\partial p_\ell} = e^i_\ell + \theta^i y_\ell(p). \tag{401}$$

Substituting (399) and (400) into (401) and rearranging yields, for $\ell < L$,

$$\frac{\partial \chi^i_{L[\psi^i]}(p)}{\partial p_\ell} = -\frac{v^i_L p_\ell}{p_L^2 \, v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} + \frac{1}{p_L}(e^i_\ell + \theta^i y_\ell(p)).$$

Differentiating (398) with respect to L yields

$$\frac{\partial \chi^i_{\ell[\psi^i]}(p)}{\partial p_L} = -\frac{1}{v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} \frac{v^i_L p_\ell}{p_L^2}$$

which in combination with (401) yields

$$\frac{\partial \chi^i_{L[\psi^i]}(p)}{\partial p_L} = \frac{v^i_L}{p_L^3} \sum_{\ell=1}^{L-1} \frac{p_\ell^2}{v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} + \frac{1}{p_L}(e^i_L + \theta^i y_L(p)). \tag{402}$$

Fixing p, choose a price-change vector $dp \not\propto p$, and let

$$\tilde{dp} \equiv dp - dp_L \, p/p_L \tag{403}$$

Observe that $\tilde{dp}_L = 0$. $D\chi^i_{[\psi^i]}(p)dp$ is thus a vector whose entries have signs opposite to those of $\tilde{dp}$. Since this holds for all i, $D\chi^i_{[\psi]}(p)dp$ also has signs opposite to those of $\tilde{dp}$.

As in Appendix B.4, $Ds(p)dp = Ds(p)\tilde{dp}$ preserves the sign of at least one nonzero entry of $\tilde{dp}$. This completes the proof that $Dz_{[\psi]}(p)dp \neq 0$ under the stated conditions.

Bounding ψ

There is a $\bar{\bar{\psi}} \gg 0$ such that, for all $(p, \psi) \in P \times \tilde{\Psi}$, if $v^i_L > \underline{v}^i_L(\psi)$ and $B^i > \underline{b}^i(\psi, v^i_L)$,

$$\psi^i(p, \psi) \in [-\bar{\bar{\psi}}, \bar{\bar{\psi}}]^L.$$

The proof proceeds precisely as in Appendix B.4. We need only let $\tilde{dp}$ notationally take the place of $\tilde{dp}_{[\psi]}$, since in this quasilinear setting the analogous adjustment to dp (defined by (403)) is independent of ψ.

Let Ψ denote the compact subset of $\tilde{\Psi}$ with $\psi^i \in [-\bar{\bar{\psi}}, \bar{\bar{\psi}}]$ $\forall i$. Since $\underline{v}^i_L(\cdot)$ is continuous, we can define

$$\underline{\underline{v}}^i_L \equiv \min_{\psi \in \Psi} \underline{v}^i_L(\psi).$$

Since $\underline{B}^i(\cdot)$ is continuous in p and ψ^i, we can define

$$\underline{\underline{B}}^i \equiv \max_{p\in P,\psi\in\Psi} \underline{B}^i(p,\psi^i,\underline{\underline{v}}^i_L).$$

Since $\underline{B}^i(\cdot)$ decreases in v^i_L, $B^i \geq \underline{\underline{B}}^i$ implies $B^i > \underline{B}^i(p,\psi^i,v^i_L) \geq \underline{b}^i(\psi,v^i_L)$ for any $p \in P$, $\psi \in \Psi$, $v^i_L > \underline{\underline{v}}^i_L$.

Here, $B^i \geq \underline{\underline{B}}^i$ and $v^i_L > \underline{\underline{v}}^i_L$ for all i, $p \in P$, and $\psi \in \Psi$ guarantee that all marginal wealth is spent on good L: $\delta(p,\psi) = a_L/p_L > 0$. So $G(p,\psi)a_L = 0$: purchases of good L do not affect prices and so do not affect supply. (Instead they only lower profits, lowering others' purchases of good L.) So $\psi^i_L(p,\psi) = 0\ \forall i$.

So, if $B^i \geq \underline{\underline{B}}^i$ and $v^i_L > \underline{\underline{v}}^i_L$ for all i,

$$\psi(p,\psi) \in \Psi \ \ \forall p \in P,\ \psi \in \Psi. \tag{404}$$

Observe that Ψ is a compact topological disc.

Existence of RCESE

Assume henceforth that $B^i \geq \underline{\underline{B}}^i$ and $v^i_L > \underline{\underline{v}}^i_L$ for all i.

In this setting, there is a WES of $(\mathcal{E}, x_{[\bar{\psi}]}(\cdot))$. It is unique up to rescaling. This follows from the fact that, for each i, $x_{[\bar{\psi}]}(\cdot)$ maximizes a quasilinear (quasi-)utility function: see Hosoya (2022), Theorem 1. Hosoya assumes that utility in goods $\ell < L$ is nondecreasing, whereas we allow marginal (quasi-)utility in such goods to be negative at sufficiently large values of x^i_ℓ, because we may have $\psi^i_\ell < -v^{i*}_\ell$. However, because each i's demands are identical to those that would obtain if marginal utility in each ℓ equaled 0 (rather than a negative number) at such large values of x^i_ℓ, Hosoya's result is maintained.

Let $p(\psi)$ denote the unique WES in P compatible with externality matrix $\psi \in \Psi$.

We will now show that $\chi_{[\psi]}(p)$ is $\mathcal{C}^1$ in both arguments throughout $P \times \Psi$. Recall that we have already established that $\chi_{[\psi]}(p)$ is $\mathcal{C}^1$ in p for any $\psi \in \Psi$. Continuity can be proven as it is surrounding (386) in the setting of Appendix B.4.

Given a marginal shift $d\psi^i$ to i's externality vector with $d\psi^i_L = 0$, holding prices and profits fixed, i's demands maintain the two optimality conditions

1. $\frac{1}{p_\ell}\frac{\partial \tilde{u}^i}{\partial x^i_\ell} = \frac{v^i_L}{p_L}$ for all ℓ,
2. $p \cdot x^i = B^i$

if they shift marginally by dx^i satisfying

$$\begin{aligned} D^2 v^i(x^i)dx^i + d\psi^i &= 0, \\ p \cdot dx^i &= 0 \\ \implies dx^i_\ell &= \frac{1}{v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} d\psi^i_\ell, \quad \ell < L; \\ dx^i_L &= \frac{1}{p_L}\sum_{\ell=1}^{L-1} \frac{p_\ell}{v^{i\prime\prime}_\ell(\chi^i_{\ell[\psi^i]}(p))} d\psi^i_\ell. \end{aligned}$$

Thus $\chi^i_{[\psi^i]}(p)$ is differentiable in ψ^i. Since $\chi^i_{\ell[\psi^i]}(p)$ is continuous in both arguments for all $\ell < L$, and $v^i_\ell(\cdot)$ is $\mathcal{C}^2$ $\forall \ell < L$, the derivative is continuous in (p, ψ^i).

The last step is to establish that the partial derivatives of $\chi^i_{[\psi^i]}(p)$ with respect to p are also continuous in (p, ψ^i) (and not just p in isolation). This follows from immediately from (399)–(402) and the continuity of $\chi^i_{[\psi]}(p)$ in both arguments.

The proof that $\psi^*(\psi) \equiv \psi(p(\psi), \psi)$ has a fixed point $\bar{\psi}$ such that $(p(\bar{\psi}), \bar{\psi})$ is an RCESE concludes as in Appendix B.4.

B.6 Proof of Proposition 9

Choose a profile of I $\mathcal{C}^1$ functions $\{w^i(\cdot)\}$ from $\mathbb{R}^L_{++}$ to $\mathbb{R}$, with $I \geq 2L+1$ and $L \geq 2$; L-vectors $\bar{s} \gg 0$, $\bar{p} \gg 0$; and an $L \times L$ matrix M with $p \in \text{Null}(M)$ and $\bar{p} \notin \text{Col}(M)$. Let

$$R \equiv \text{Rank}(M).$$

Define ψ by $\psi^i = M\nabla w^i(\bar{s})\ \forall i$.

Choose a positive semidefinite, symmetric, $L \times L$ matrix N with $\text{Col}(N) = \text{Col}(M^T)$ (and thus $\text{Rank}(N) = R$), $N\bar{p} = \mathbb{0}_L$, and $\bar{p} \cdot N = \mathbb{0}_L^T$.

There is a (not necessarily unique) matrix G such that $NG = M^T$, $\text{Rank}(G) = L-1$, and $G_{L\ell} = 0\ \forall \ell$, i.e. the bottom row of G consists of zeroes. To construct one, let $m_\ell \equiv M^T a_\ell$ denote the ℓ^{th} column of M^T. Since $m_\ell \in \text{Col}(M^T)$, $m_\ell \in \text{Col}(N)$. It follows that, for each ℓ, there is a vector $\tilde{g}_\ell$ with

$$N\tilde{g}_\ell = m_\ell.$$

Choose $\{\tilde{g}_\ell\}_{\ell=1}^L$. Each can be decomposed into the sum of an element of $\text{Null}(N)$ and a vector $\tilde{\tilde{g}}_\ell$ orthogonal to $\text{Null}(N)$. So $\{\tilde{\tilde{g}}_\ell\}_{\ell=1}^L$ is a set of vectors orthogonal to $\text{Null}(N)$ such that

$$N\tilde{\tilde{g}}_\ell = m_\ell\ \ \forall \ell.$$

R of these vectors are linearly independent; if fewer, then $\text{Rank}(G) < R$ and thus $\text{Rank}(M) < R$. Choose R linearly independent elements of $\{\tilde{\tilde{g}}_\ell\}$, and for these, define $\tilde{\tilde{\tilde{g}}}_\ell = \tilde{\tilde{g}}_\ell$. For each of the $L - R$ elements remaining, define $\tilde{\tilde{\tilde{g}}}_\ell$ by adding a distinct element of a basis of the $(L-R)$-dimensional space $\text{Null}(N)$ to $\tilde{\tilde{g}}_\ell$. We now have a

set of linearly independent vectors $\tilde{\tilde{g}}_\ell$ with

$$N\tilde{\tilde{g}}_\ell = m_\ell \ \ \forall \ell.$$

Construct G by setting its ℓ^{th} column equal to $g_\ell \equiv \tilde{\tilde{g}}_\ell - [\tilde{\tilde{g}}_\ell]_L/\bar{p}_L\,\bar{p}$. By construction, $G_{L\ell} = 0\ \forall \ell$. Since $\bar{p} \in \text{Null}(N)$, $Ng_\ell = m_\ell\ \forall \ell$; $NG = M^T$.

Since $Ga_L = 0$, $\text{Rank}(G) < L$. To verify that $\text{Rank}(G) \geq L-1$, observe that

$$G = \tilde{\tilde{G}} - \bar{p}\gamma^T,$$

where $\tilde{\tilde{G}}$ is the $L \times L$ matrix with column ℓ equaling $\{\tilde{\tilde{g}}_\ell\}$ and γ is the L-dimensional column vector with $\gamma_\ell = [\tilde{\tilde{g}}_\ell]_L/\bar{p}_L$. Because $\tilde{\tilde{G}}$ is of full rank, there is a vector $\tilde{\gamma}$ with

$$\tilde{\tilde{G}}\tilde{\gamma} = \bar{p},$$

and thus

$$G = \tilde{\tilde{G}}(I_L - \tilde{\gamma}\gamma^T),$$

and thus $\text{Rank}(G) = \text{Rank}(I_L - \tilde{\gamma}\gamma^T)$. If $\text{Rank}(I_L - \tilde{\gamma}\gamma^T) < L-1$, there are linearly independent vectors b, c with $\tilde{\gamma}\gamma^T b = b$ and $\tilde{\gamma}\gamma^T c = c$. This is impossible, since the columns of $\tilde{\gamma}\gamma^T$ are all proportional to $\tilde{\gamma}$, so $\text{Rank}(\tilde{\gamma}\gamma^T) \leq 1$. This completes the proof that $\text{Rank}(G) = L-1$.

We will now construct a matrix G^* with $G^*Gh = h$ for $h \perp \bar{p}$ and $G^*\bar{p} = 0$.

Observe that $\text{Col}(G) = \{b \in \mathbb{R}^L : b_L = 0\}$. Choose vectors $\{\tilde{d}_\ell\}_{\ell=1}^{L-1}$ that satisfy $G\tilde{d}_\ell = a_\ell$, and let

$$d_\ell \equiv \tilde{d}_\ell - (\tilde{d}_\ell \cdot \bar{p})\delta \ \ \forall \ell.$$

Because $d \cdot \bar{p} = 1$, $d_\ell \cdot \bar{p} = 0$; $d_\ell \perp \bar{p}\ \forall \ell$. Also, because $Gd_\ell = a_\ell\ \forall \ell$, the set $\{d_\ell\}$ is linearly independent. So $\{d_\ell\}$ spans the hyperplane orthogonal to $\bar{p}$.

If G^* is an $L \times L$ matrix whose ℓ^{th} column is d_ℓ for $\ell < L$, then by construction, $G^*Gh = h$ for $h \perp \bar{p}$. $G^*\bar{p} = 0$ is then be maintained by setting $G^*_{\ell L} = -\sum_{k=1}^{L-1} G_{\ell k}\bar{p}_k/\bar{p}_L$ for each ℓ.

By construction, we also have $\bar{p} \cdot G^* = \mathbb{0}_L^T$.

Choose a nonzero vector $\delta \in \text{Null}(G) \subset \text{Null}(M^T)$. Since $\bar{p} \notin \text{Col}(M)$ and the left null space is orthogonal to the column space, $\bar{p} \cdot \delta \neq 0$. Without loss of generality, choose δ such that $\bar{p} \cdot \delta = 1$.

Choose a production function $y(\cdot)$ that is $\mathcal{C}^1$ around $\bar{p}$, with $y(\bar{p}) = 0$ and $Dy(\bar{p}) = N$.

Choose an aggregate demand function $x(\cdot)$ that is $\mathcal{C}^2$ around $(\bar{p}, 0)$, with $x(\bar{p}, 0) = \bar{s}$ and $D_p x(\bar{p}, 0) = Ds(\bar{p}) - G^*$.

Choose $\bar{x}^I \gg 0, \ll \bar{s}$ and let $\bar{x}^{-I} \equiv \bar{s} - \bar{x}^I$. Let $\theta^I = 1$. Choose an increasing, strictly concave consumption utility function $v^I(\cdot)$ and an individual endowment e^I such that

- $x^I_{[0]}(\bar{p}, 0)\ (\equiv \tilde{x}^I_{[0]}(\bar{p}, \bar{p} \cdot e^I)) = e^I$,
- $x^I_{[0]}(p, \pi)\ (\equiv \tilde{x}^I_{[0]}(p, \bar{p} \cdot e^I + \pi))$ is defined and $\mathcal{C}^1$ in (p, π) in a neighborhood of $(\bar{p}, 0)$ (regarding $\tilde{x}^I_{[0]}(\cdot)$: $\mathcal{C}^1$ in p, B^i in a neighborhood of $(\bar{p}, \bar{p} \cdot e^I)$), and
- $\nabla_\pi x^I_{[0]}(\bar{p}, 0)\ (\equiv \nabla_{B^I} \tilde{x}^I_{[0]}(\bar{p}, \bar{p} \cdot e^I)) = \delta$.

The wealth effect vector at specified prices $\bar{p}$ (here δ) and the Jacobian of demand with respect to prices at $\bar{p}$ can be set independently, subject to $\bar{p} \cdot \delta = 1$; see Mas-Colell et al. (1995), pp. 600–601.

Let $z^*(p) \equiv x(p,0) - \bar{s}$ denote the excess demand function generated by aggregate demand function $x(\cdot)$ in an exchange economy with aggregate endowment vector $\bar{s}$. It is also the excess demand function generated by aggregate demand function $x(\cdot) - e^I$ in an exchange economy with aggregate endowment vector $\bar{s} - e^I$. Because $x(\cdot) - e^I$ is $\mathcal{C}^1$ in p around $(\bar{p}, 0)$, there is an $\epsilon > 0$ such that, within $\mathcal{N}_\epsilon(\bar{p})$, $z^*(\cdot)$ satisfies the second-to-last inequality near the bottom of p. 349 of Mantel (1974), with our p, $\bar{p}$, and $z^*_\ell(\cdot)$ (for each ℓ) taking the place of Mantel's q, p, and $g(\cdot)$.

Then by the theorem of Mantel (1974), there is a profile of $I-1$ continuous, increasing, strictly concave consumption utility functions $\{v^i(\cdot)\}$ and endowments $\{e^i\}$ with $e = \bar{s} - e^I$ implying the following demand functions, written as functions of prices and wealth:

$$\begin{aligned}
\tilde{x}^i_{[0]}(p, B^i) &= \frac{B^i}{L}\mathrm{Diag}(p)^{-1}\mathbb{1}_L - \frac{B^i}{\kappa^i}Dz^*(p)a_i, && i = 1, ..., L; \\
&= \frac{\kappa^i}{L}\big(\mathbb{1}_L - p \cdot \mathbb{1}_L\mathrm{Diag}(p)^{-1}a_{i-L}\big) + \frac{B^i}{p_{i-L}}a_{i-L}, && i = L+1, ..., I-1
\end{aligned}$$

for some nonzero constants $\{\kappa^i\}$, such that

$$\sum_{i=1}^{I-1} \tilde{x}^i_{[0]}(p, p \cdot e^i) - \bar{s} = z^*(p) \text{ for } p \in \mathcal{N}_\epsilon(\bar{p}).$$

(Mantel notes only that his implied utility functions are monotone—i.e. weakly increasing—rather than strictly increasing; but because they are strictly quasiconcave throughout $\mathbb{R}^L_+$ (not just $\mathbb{R}^L_{++}$), they cannot be constant in any good throughout $\mathbb{R}^L_+$. Also, Mantel's construction consists of precisely $2L$ consumers; but given a construction for ν consumers, it can be extended to include $\nu + 1$ consumers by "dividing" some i into collectively behaviorally equivalent j_1, j_2 with $e^j = e^i/2$ and $v^j(x^j) = v^i(2x^j)$ for $j \in \{j_1, j_2\}$.)

Given these endowments and profit shares $\theta^i = 0$ for $i < I$, we can write these individual demands as functions of prices and (trivially) profits:

$$x^i_{[0]}(p, \pi) = \frac{p \cdot e^i}{L} \text{Diag}(p)^{-1} \mathbb{1}_L - \frac{p \cdot e^i}{\kappa^i} Dz^*(p) a_\ell.$$

Since $z^*(p)$ is $\mathcal{C}^2$, $x^i_{[0]}(\cdot)$ is $\mathcal{C}^1$ in p, π for all $i < I$.

The continuous differentiability of $x_{[\mathbb{0}_{L\times I}]}(\cdot)$ and $y(\cdot)$, $\bar{p} \cdot \delta > 0$, and the implicit function theorem imply that the profit function $\pi(p, x_{[\mathbb{0}_{L\times I}]}(\cdot))$ is $\mathcal{C}^1$ in p around $\bar{p}$. So $\chi_{[\mathbb{0}_{L\times I}]}(\cdot)$ is $\mathcal{C}^1$ in p around $\bar{p}$. Also, $D\chi_{[\mathbb{0}_{L\times I}]}(\bar{p}) = D_p x(\bar{p}, 0)$, given $y(\bar{p}) = 0$ and the envelope theorem (which implies $\nabla_p \pi(p, x_{[\mathbb{0}_{L\times I}]}(\cdot)) = y(p)$). Summing the individual demands, $\chi_{[\mathbb{0}_{L\times I}]}(\bar{p}) = \bar{s} - e^I + e^I = \bar{s}$.

We now have an environment $(\{e^i\}, \{\theta^i\}, y(\cdot))$ and profile of consumption utility functions $\{v^i(\cdot)\}$ such that $s(\bar{p}) = \bar{s}$ and $Ds(\bar{p})G(\bar{p}, x_{[\mathbb{0}_{L\times I}]}(\cdot)) = M^T$. The latter equality holds because $Ds(\bar{p})G = M^T$, where by construction G equals the unique generalized inverse of $G^* = Ds(\bar{p}) - D\chi_{[\mathbb{0}_{L\times I}]}(\bar{p})$ with $G\nabla_\pi x_{[\mathbb{0}_{L\times I}]}(p, \pi) = G\delta = 0$ and a bottom row of zeroes; i.e. $G = G(\bar{p}, x_{[\mathbb{0}_{L\times I}]}(\cdot))$. It follows that

$$\psi(\bar{p}, x_{[\mathbb{0}_{L\times I}]}) = \psi.$$

Since for each i $v^i(\cdot)$ is strictly concave and increasing in all goods, its right derivative in each good ℓ (which we will denote $v^{i\prime}_\ell(\cdot)$) is everywhere defined and positive. Let $\mathcal{B}^i(p)$ denote i's budget set at prices $p \gg 0$, and choose an $\epsilon > 0$ and let

$$\mathcal{B}^i \equiv \{x^i : \bar{p} \cdot x^i \leq \bar{p} \cdot e^i + \epsilon\} \tag{405}$$

denote the budget set i faces given prices $\bar{p}$ under a relaxed budget constraint. Ob-

serve that $\mathcal{B}^i(p) \subset \mathcal{B}^i$ for all prices p near $\bar{p}$.

We will now show that for each i, for all $\bar{x}^i \in \mathcal{B}^i$, there is an $\epsilon^i(\bar{x}^i) > 0$ such that, for all $x^i \in \mathcal{N}_{\epsilon^i(\bar{x}^i)}(\bar{x}^i)$, $v^{i\prime}_\ell(x^i) > \underline{v}^{i\prime}_\ell$.

Let $\overline{\mathcal{B}}^i$ denote the upper boundary of $\mathcal{B}^i$: i.e. (405) with the weak inequality replaced with an equality. Fix i and $\bar{x}^i \in \overline{\mathcal{B}}^i$. The continuity and strict monotonicity of $v^i(\cdot)$ imply that there is an $r \gg 0$ and an indifference surface "IS_1", corresponding to consumption utility level $\bar{v}_1 > v^i(\bar{x}^i)$, that passes through $\bar{x}^i + r_\ell a_\ell$ for all ℓ. These properties also imply that IS_1 is bounded away from $\bar{x}^i$. There are thus also indifference surfaces "IS_2" and "IS_3", both bounded away from $\bar{x}^i$, corresponding to consumption utility levels $\bar{v}_2$, $\bar{v}_3$ with $v^i(\bar{x}^i) < \bar{v}_3 < \bar{v}_2 < \bar{v}_1$.

For some $\epsilon > 0$, there is a $\bar{r} \gg 0$ with $v^i(\bar{x}^i(1-\epsilon) + \bar{r}_\ell a_\ell) = \bar{v}_2$ for all ℓ. If for all $\epsilon > 0$ there were no such $\bar{r}$, then (by continuity and monotonicity of $v^i(\cdot)$) there would be an n-indexed sequence $\bar{x}^i(1-1/n) \to \bar{x}^i$ and an ℓ such that $v^i(\bar{x}^i(1-1/n) + r_\ell a_\ell) < \bar{v}_2$ for all n, even though by construction in the limit we have $v^i(\bar{x}^i(1-1/n) + r_\ell a_\ell) = \bar{v}_1 > \bar{v}_2$ (contradicting the continuity of $v^i(\cdot)$). Observe also that we must have $\bar{r} \gg \epsilon \bar{x}^i$, by monotonicity of $v^i(\cdot)$ and $\bar{v}_2 > v^i(\bar{x}^i)$. Choose such an ϵ and $\bar{r}$.

Let

$$\underline{x}^i \equiv \bar{x}^i(1-\epsilon)$$

and

$$X \equiv \bigtimes_{\ell=1}^{L} \left[\underline{x}^i_\ell,\ \underline{x}^i_\ell + \bar{r}_\ell\right].$$

That is, let X denote the box ranging from $\underline{x}^i_\ell$ to $\underline{x}^i_\ell + \bar{r}_\ell$ for each dimension ℓ. Observe that, given any point $x^i \in \text{IS}_3 \cap X$ and any ℓ, there is a unique and positive scalar $r_\ell(x^i) < \bar{r}_\ell$ such that $v^i(x^i + r_\ell(x^i)a_\ell) = \bar{v}_2$. (Existence follows from the continuity of $v^i(\cdot)$: consumption utility level $\bar{v}_2$ must be achievable by beginning

with x^i and adding a quantity of good ℓ, since by monotonicity and $x^i > \underline{x}^i$ we have $v^i(x^i + \bar{r}_\ell a_\ell) > v^i(\underline{x}^i + \bar{r}_\ell a_\ell) = \bar{v}_2$.) By strict concavity, it follows that the right derivative of $v^i(\cdot)$ in good ℓ at x^i is greater than

$$\underline{v}_\ell^{i\prime}(\bar{x}^i) \equiv (\bar{v}_2 - \bar{v}_3)/\bar{r}_\ell > 0.$$

Then likewise, because IS_3 is bounded away from $\bar{x}^i$, there is an $\epsilon(\bar{x}^i) > 0$ such that $\mathcal{N}_{\epsilon^i(\bar{x}^i)}(\bar{x}^i) \cap \mathbb{R}^L_+ \subset X$ and, for all ℓ and $x^i \in \mathcal{N}_{\epsilon(\bar{x}^i)}(\bar{x}^i) \cap \mathbb{R}^L_+$, we have $x^i + \bar{r}_\ell(x^i) a_\ell \in \text{IS}_3 \cap X$ for some $\tilde{r}_\ell(x^i) > 0$. By strict concavity, the right derivative of $v^i(\cdot)$ in good ℓ at this x^i is greater than $\underline{v}_\ell^{i\prime}(\bar{x}^i)$.

The set $\{\mathcal{N}_{\epsilon^i(\bar{x}^i)}(\bar{x}^i)\}_{\bar{x}^i \in \overline{\mathcal{B}}^i}$ constitutes an open cover of $\overline{\mathcal{B}}^i$. Because $\overline{\mathcal{B}}^i$ is compact, we can choose a finite subcover. There is thus for each ℓ a $\underline{v}_\ell^{i\prime} > 0$ such that the right derivative of $v^i(\cdot)$ in good ℓ at $\bar{x}^i$ is at least $\underline{v}_\ell^{i\prime}$ for all $\bar{x}^i \in \overline{\mathcal{B}}^i$.

Finally, by strict concavity, the $\{\underline{v}_\ell^{i\prime}\}$ lower-bound the right derivatives throughout $\mathcal{B}^i$.

Consider the consumption utility function profile $\{\tilde{v}^i(\cdot)\}$ with

$$\tilde{v}^i(x^i) = K^i v^i(x^i) - \psi^i \cdot x^i,$$

where

$$K^i = \max\Big(1,\ \max_\ell\ -2\frac{\psi_\ell^i}{\underline{v}_\ell^{i\prime}}\Big).$$

By construction, the right derivative of $\tilde{v}^i(\cdot)$ in each ℓ is positive throughout $\mathcal{B}^i$ for each i. Observe also that $\tilde{v}^i(\cdot)$ inherits the strict concavity of $v^i(\cdot)$: given any

$\alpha \in [0,1]$, x^i, and $\tilde{x}^i$, we have

$$\begin{aligned}
&K^i v^i(\alpha x^i + (1-\alpha)\tilde{x}^i) - \psi^i \cdot (\alpha x^i + (1-\alpha)\tilde{x}^i) \\
&> \alpha K^i v^i(x^i) + (1-\alpha) K^i v^i(\tilde{x}^i) - \alpha \psi^i \cdot x^i - (1-\alpha)\psi^i \cdot \tilde{x}^i \\
&= \alpha\big(K^i v^i(x^i) - \psi^i \cdot x^i\big) + (1-\alpha)\big(K^i v^i(\tilde{x}^i) - \psi^i \cdot \tilde{x}^i\big).
\end{aligned}$$

Let $\tilde{\tilde{v}}^i(x^i) = \tilde{v}^i(x^i)$ for $x^i \in \mathcal{B}^i$ and let $\tilde{\tilde{v}}^i(\cdot)$ be increasing and strictly concave in all goods above the budget constraint, so that $\tilde{\tilde{v}}^i(\cdot)$ is globally increasing and strictly concave. Given consumption utility function profile $\{\tilde{\tilde{v}}^i(\cdot)\}$ and externality matrix ψ, for all i we have demands $\chi^i_{[\psi]}(p)$ equal to the quantities expressed above as $\chi^i_{[\mathbb{0}_{L\times I}]}(p)$ (with respect to $\{v^i(\cdot)\}$) throughout some neighborhood of $\bar{p}$. It follows that, given utility function profile $\{\tilde{\tilde{v}}^i(\cdot) + w^i(\cdot)\}$, $\psi(\bar{p}, \psi) = \psi$.

References

Admati, Anat R. and Motty Perry, "Joint Projects without Commitment," *Review of Economic Studies*, 1991, *58* (2), 259–276.

Alan, Sule and Martin Browning, "Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation," *Review of Economic Studies*, 2010, *77* (4), 1231–1261.

Andreessen, Marc, "The Techno-Optimist Manifesto," https://a16z.com/the-techno-optimist-manifesto/ (accessed 2 Oct 2024), October 2023. Open letter.

Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 1990, *100* (1), 464–477.

Aumann, Robert J., "Markets with a Continuum of Traders," *Econometrica*, 1964, *32* (1–2), 39–50.

Aurland-Bredesen, Kine Josefine, "The Optimal Economic Management of Catastrophic Risk." PhD dissertation, Norwegian University of Life Sciences School of Economics and Business 2019.

Bade, Sophie, "Nash Equilibrium in Games with Incomplete Preferences," *Economic Theory*, 2005, *26*, 309–332.

Baranzini, Andrea and François Bourguinion, "Is Sustainable Growth Optimal?," *International Tax and Public Finance*, 1995, *2*, 341–356.

Barrage, Lint, "Be Careful What You Calibrate For: Social Discounting in General Equilibrium," *Journal of Public Economics*, 2018, *160*, 33–49.

Barro, Robert J. and Xavier I. Sala-i-Martin, *Economic Growth, Second Edition*, Cambridge, MA: The MIT Press, 2004.

Battaglini, Marco, Salvatore Nunnari, and Thomas R. Palfrey, "Dynamic Free Riding with Irreversible Investments," *American Economic Review*, 2014, *104* (9), 2858–2871.

Becker, Robert A., "On the Long-run Steady State in a Simple Dynamic Model of Equilibrium with Heterogeneous Households," *Quarterly Journal of Economics*, 1980, *95* (2), 375–382.

Belfiori, Maria E., "Carbon Pricing, Carbon Sequestration, and Social Discounting," *European Economic Review*, 2017, *96*, 1–17.

Benhabib, Jess and Roy Radner, "The Joint Exploitation of a Productive Asset: A Game-theoretic Approach," *Economic Theory*, 1992, *2*, 155–190.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian, "On the Private Provision of Public Goods," *Journal of Public Economics*, 1986, *29* (1), 25–49.

Bostrom, Nick, "Existential Risks: Analyzing Human Extinction Scenarios," *Journal of Evolution and Technology*, March 2002, *9* (1), 1–35.

_ , *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014.

_ , "The Vulnerable World Hypothesis," *Global Policy*, 2019, *10* (4), 455–476.

Brock, William A. and M. Scott Taylor, "Economic Growth and the Environment: A Review of Theory and Empirics," in Philippe Aghion and Steven N. Durlauf, eds., *The Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, chapter 28, pp. 1749–1821.

Budolfson, Mark Bryant, "The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response," *Philosophical Studies*, 2019, *176* (7), 1711–1724.

Caputo, Michael R., *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*, Cambridge University Press, 2005.

Chambers, Christopher P. and Federico Echenique, "On Multiple Discount Rates," *Econometrica*, 2018, *86* (4), 1325–1346.

Chetty, Raj, "A Bound on Risk Aversion Using Labor Supply Elasticities," *American Economic Review*, 2006, *96* (5).

Christiano, Paul, "Analyzing divestment," https://sideways-view.com/2019/05/25/analyzing-divestment/ (accessed 2 Oct 2024), 2019. "The Sideways View", blog post.

Cowen, Tyler and Derek Parfit, "Against the Social Discount Rate," in Peter Laslett and James S. Fishkin, eds., *Justice Between Age Groups and Generations*, New Haven: Yale University Press, 1992, pp. 144–161.

Drupp, Moritz, Mark Freeman, Ben Groom, and Frikk Nesje, "Discounting Disentangled," *American Economic Journal: Economic Policy*, 2018, *10* (4), 109–134.

Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel, "Other-Regarding Preferences in General Equilibrium," *Review of Economic Studies*, 2011, *78* (2), 613–39.

Dutta, Prajit, "Dynamic Games with an Application to Climate Change Models," in Robert A. Myers, ed., *Encyclopedia of Complexity and Systems Science*, New York: Springer, 2017.

Ellickson, Bryan, *Competitive Equilibrium: Theory and Applications*, Cambridge, UK: Cambridge University Press, 1994.

Elsey, Jamie and David Moss, "Time preferences for impact in the EA community (data from the 2023 EA Survey Supplement)," https://forum.effectivealtruism.org/posts/DncphHoNnmKj685ya/time-preferences-for-impact-in-the-ea-community-data-from (accessed 12 Aug 2024), August 2024. Effective Altruism Forum.

Farhi, Emmanuel and Ivan Werning, "Inequality and Social Discounting," *Journal of Political Economy*, 2007, *115* (3), 365–402.

_ **and** _ , "Progressive Estate Taxation," *Quarterly Journal of Economics*, 2010, *125* (2), 635–673.

Farquhar, Sebastian, John Halstead, and Owen Cotton-Barratt, "Existential Risk: Diplomacy and Governance," Technical Report 2017.

Fershtman, Chaim and Shmuel Nitzan, "Dynamic Voluntary Provision of Public Goods," *European Economic Review*, 1991, *35*, 1057–1067.

Fishman, Ram, "Heterogeneous Patience, Bargaining Power and Investment in Future Public Goods," *Environmental and Resource Economics*, 2019, *73*, 1101–1107.

Frietas-Groff, Zachary and Karan Makkar, "Positive Social Discount Rates: An Experiment on Equity Preferences Across Cohorts," 2023. Working paper.

Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," March 2023.

Gardner, Bruce L., "Income Distribution through Multi-market Commodity Programs," in Bruce L. Gardner, ed., *The Economics of Agricultural Policies*, New York: Macmillan, 1987.

Gollier, Christian and Richard Zeckhauser, "Aggregation of Heterogeneous Time Preferences," *Journal of Political Economy*, 2005, *113* (4), 878–96.

Green, Daniel and Benjamin Roth, "The Allocation of Socially Responsible Capital," *Journal of Finance*, 2024. Forthcoming.

Hall, Robert, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 1988, *96* (2), 339–357.

Heal, Geoffrey M. and Antony Millner, "Agreeing to Disagree on Climate Policy," *Proceedings of the National Academy of Sciences*, 2014, *111* (10), 3695–3698.

HM Treasury, "The Green Book: Central Government Guidance on Appraisal and Evaluation," https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938046/The_Green_Book_2020.pdf (accessed 10 Feb 2020), 2020.

Hosoya, Yuhki, "On the Uniqueness and Stability of the Equilibrium Price in Quasi-Linear Economies," 2022. Working paper.

Isaacs, Yoaav, Adam Lerner, and Jeffrey Sanford Russell, "Counting Your Chickens," *Australasian Journal of Philosophy*, 2024, *102* (3).

Jackson, Matthew and Alejandro Manelli, "Approximately Competitive Equilibria in Large Finite Economies," *Journal of Economic Theory*, 1997, *77* (2), 354–376.

Jackson, Matthew O. and Leeat Yariv, "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 2015, *7* (4), 150–178.

Jacobsen, Mark, Jacob LaRiviere, and Michael Price, "Public Policy and the Private Provision of Public Goods under Heterogeneous Preferences," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (1), 243–280.

Jones, Charles I., "Life and Growth," *Journal of Political Economy*, 2016, *124* (2), 539–578.

_ , "The A.I. Dilemma: Growth Versus Existential Risk," 2024. Working paper.

Kalai, Adam and Ehud Kalai, "Strategic Polarization," *Journal of Mathematical Psychology*, 2001, *45*, 656–663.

Kaufmann, Marc, Peter Andre, and Botond Kőszegi, "Understanding Markets with Socially Responsible Consumers," *Quarterly Journal of Economics*, 2024, *139* (3), 1989–2035.

Klenow, Peter J., Charles I. Jones, Mark Bils, and Mohamad Adhami, "Population and Welfare: The Greatest Good for the Greatest Number," June 2024.

Konovalov, Alexander, "The Core of an Economy with Satiation," *Economic Theory*, 2005, *25*, 711–719.

Kreps, David M., *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press, 1990.

_ , *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton: Princeton University Press, 2012.

Li, Qingran and William A. Pizer, "Discounting for Public Benefit-Cost Analysis," 2021. Resources for the Future, Issue Brief 21-05.

Lucas, Deborah, "Asset Pricing with Undiversifiable Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle," *Journal of Monetary Economics*, 1994, *34* (3), 325–342.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22*, 3–42.

MacAskill, William, *What We Owe the Future: A Million-Year View*, Oneworld Publications, 2022.

Mantel, Rolf, "On the Characterization of Aggregate Excess Demand," *Journal of Economic Theory*, 1974, *7* (3), 348–53.

Martin, Ian W. R. and Robert S. Pindyck, "Averting Catastrophes: The Strange Economics of Scylla and Charybdis," *American Economic Review*, 2015, *105* (10), 2947–2985.

_ **and** _ , "Welfare Costs of Catastrophes: Lost Consumption and Lost Lives," *The Economic Journal*, 2021, *131* (634), 946–969.

Marx, Leslie M. and Steven A. Matthews, "Dynamic Voluntary Contribution to a Public Project," *Review of Economic Studies*, 2000, *67* (2), 327–358.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Theory*, Oxford: Oxford University Press, 1995.

McMillan, John, "Individual Incentives in the Supply of Public Inputs," *Journal of Public Economics*, 1979, *12*, 87–98.

Meadows, Donella H., Dennis L. Meadows, Jørgen Randers, and William W. Behrens III, *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*, New York: Universe Books, 1972.

Millett, Piers and Andrew Snyder-Beattie, "Existential Risk and Cost-Effective Biosecurity," *Health Security*, 2017, *15*, 373–383.

Millner, Antony, "Nondogmatic Social Discounting," *American Economic Review*, 2020, *110* (3), 760–775.

_ **and Geoffrey Heal**, "Discounting by Committee," *Journal of Public Economics*, 2018, *167*, 91–104.

Moynihan, Thomas, *X-Risk: How Humanity Discovered Its Own Extinction*, Urbanomic, 2020.

National Philanthropic Trust, "The 2020 DAF Report," https://www.nptrust.org/reports/daf-report/ (accessed 30 Jun 2020), 2020.

Nesje, Frikk, *Journal of Environmental Economics and Management*, 2024, *128.*

Nordhaus, William, "A Review of the *Stern Review on the Economics of Climate Change*," *Journal of Economic Literature*, 2007, *45* (3), 686–702.

Norwood, F. Bailey and Jayson L. Lusk, *Compassion, by the Pound: The Economics of Farm Animal Welfare*, Oxford: Oxford University Press, 2011.

O'Kane, Lawrence, "U.S. Fights Trusts That Provide 500 Years of Compound Interest," *New York Times*, 7 1961.

Ord, Toby, *The Precipice: Existential Risk and the Future of Humanity*, New York: Bloomsbury, 2020.

_ , "Robust longterm comparisons," https://www.tobyord.com/writing/robust-longterm-comparisons (accessed 2 Oct 2024), 2024. Blog post.

Otani, Yoshihiko and Joseph Sicilian, "Limit Properties of Equilibrium Allocations of Walrasian Strategic Games," *Journal of Economic Theory*, 1990, *51* (2), 295–312.

Parfit, Derek, *Reasons and Persons*, Oxford: Oxford University Press, 1984.

Pastor, Lubos, Robert Stambaugh, and Lucian A. Taylor, "Sustainable Investing in Equilibrium," *Journal of Financial Economics*, 2021, *142* (2), 550–571.

Posner, Richard A., *Catastrophe: Risk and Response*, New York: Oxford University Press, 2004.

Rader, Trout, "Utility over Time: The Homothetic Case," *Journal of Economic Theory*, 1981, *25* (2), 219–236.

Roberts, John and Andrew Postlewaite, "The Incentives for Price-taking Behavior in Large Exchange Economies," *Econometrica*, 1976, *44* (1), 115—-27.

Roth, Benjamin, "Impact Investing: A Theory of Financing Social Enterprises," 2021. Harvard Business School, Working Paper 20-078.

Roth Tran, Brigitte, "Divest, Disregard, or Double Down? Philanthropic Endowment Investments in Objectionable Firms," *American Economic Review: Insights*, 2019, *1* (2), 241–256.

Ryder, Harl, "Heterogeneous Time Preferences and the Distribution of Wealth," *Mathematical Social Sciences*, 1985, *9* (1), 63–76.

Sagan, Carl, *Pale Blue Dot: A Vision of the Human Future in Space*, Ballantine Books, 1997.

Shulman, Carl and Elliott Thornley, "How Much Should Governments Pay to Prevent Catastrophes? Longtermism's Limited Role," in Jacob Barrett, Hilary Greaves, and David Thorstad, eds., *Essays on Longtermism*, Oxford: Oxford University Press, 2024.

Snyder-Beattie, Andrew E., Toby Ord, and Michael B. Bonsall, "An Upper Bound for the Background Rate of Human Extinction," *Scientific Reports*, December 2019, *9* (1), 11054.

Sobel, Joel, "Generous Actors, Selfish Actions: Markets with Other-Regarding Preferences," *International Review of Economics*, 2009, *56*, 3–16.

Sorger, Gerhard, "Recursive Nash Bargaining over a Productive Asset," *Journal of Economic Dynamics and Control*, 2006, *30* (12), 2637–2659.

Stern, Nicholas, *The Economics of Climate Change: The Stern Review*, Cambridge and New York: Cambridge University Press, 2007.

Stinchcombe, Maxwell B., "Continuous Time Game Theory: An Infinitesimal Approach," http://www.laits.utexas.edu/~mbs31415/InfGTc.pdf (accessed 23 Sep 2020), 2013.

Stokey, Nancy, "Are There Limits to Growth?," *International Economic Review*, 1998, *39* (1), 1–31.

US Office of Management and Budget, "Circular A-94: Guidelines and Discount Rates for Benefit-Cost Analysis Of Federal Programs," https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A94/a094.pdf (accessed 10 Feb 2020), 2019.

von Below, David, "Optimal Carbon Taxes with Social and Private Discounting," 2012. Working Paper, SURED Conference (Ascona, Switzerland).

Wilkinson, Hayden, "Market Harms and Market Benefits," *Philosophy & Public Affairs*, 2022. Online preprint.

Wohlgenant, Michael K., "Consumer Demand and Welfare in Equilibrium Displacement Models," in Jayson L. Lusk, Jutta Roosen, and Jason F. Shogren, eds., *The Oxford Handbook of the Economics of Food Consumption and Policy*, Oxford: Oxford University Press, 2011.