Efficiency and Endogenous Growth

Philip Trammell*

May 4, 2025

Abstract

Since 1998, endogenous growth theory has reconciled rising populations with constant growth rates by positing that, as population grows, research effort is spread over an ever wider range of goods. As usually modeled, this implies that decentralized and historical growth are so inefficient that infinite utility would be feasible, with a fixed or growing population, by focusing research effort on ever fewer goods. This result is avoided only by positing sublogarithmic returns to research inputs. This has implications for the plausibility of endogenous growth theory in its current form, as well as for the growth impact of automation.

1 Introduction

Two schools in modern growth theory — Output per person has long grown roughly exponentially throughout the industrialized world. Romer (1990) introduced the idea that this is because output Y per person L is governed in the long run by the quantity of nonrival technology A,

$$Y_t = A_t(1 - S_t)L_t,$$

where A grows according to[1]

$$\dot{A}_t \propto A_t S_t L_t,$$

and S denotes the fraction of the population engaged in research. (Everyone works.) The idea is that technology assists the population of researchers SL in the process of technological development roughly as it assists producers in production. Since S is bounded

*Digital Economy Lab, Stanford University. Contact: trammell@stanford.edu. Thanks to Tom Davidson for inspiring the idea, by making this point about Peretto (2018), and to Chad Jones for encouraging me to write the paper.

[1] The symbol ∝ means "proportional to".

above by 1, its growth cannot sustain growth in technology, so let us assume for simplicity that it is fixed. Let us also generalize from the assumption that the "research function" governing $\dot{A}$ is linear in research effort to the assumption that it is merely increasing:

$$Y_t \propto A_t L_t, \quad \dot{A}_t \propto A_t R(L_t) \text{ with } R' > 0.$$

The upshot is that the [economic] *growth rate* $\dot{A}_t/A_t$ is constant if the population is constant but increases in the population size. Other early endogenous growth models (Grossman and Helpman (1991), Aghion and Howitt (1992)) share this property.

But populations have grown greatly over the last century (and indeed research populations have grown even more quickly). Modern growth theory offers two approaches to reconciling this fact with constant growth.

Semi-endogenous growth models (Jones (1995), Kortum (1997), Segerstrom (1998)) posit

$$\dot{A}_t \propto A_t^{\phi} L_t^{\lambda} \text{ for } \phi < 1, \lambda > 0,$$

or some minor variation on the above.[2] Thus growth is power-functional if the population is fixed, and exponential, in steady state, only if the population grows exponentially. These are not the subject of this paper, except insofar as it contributes to the long-standing debate over whether semi-endogenous models are more or less plausible than the alternative.

Second-wave endogenous (I will write "SWE") growth models posit instead that technology grows exponentially with a constant or with a growing population. The idea, with minor variations, is that process efficiency—the quantity of a given good[3] producible with given labor and/or capital inputs—grows exponentially with constant research inputs, as in a first-wave model; but when the population doubles, we develop twice as many goods, leaving research inputs per good fixed. Improvements in process efficiency are called "vertical innovations"; increases in good variety are called "horizontal innovations". Variety may or may not be desirable, so an increase to the population size may or may not yield a level effect, but in either case it does not yield a growth effect. Likewise exponential population growth may raise the technology growth rate from the positive rate that obtains when the population is fixed, by adding a horizontal dimension to the constant vertical dimension, but even if so, the result is still a constant rather than a rising growth rate. Young (1998) introduced the SWE approach, and its latest iteration is as recent as Aghion et al. (2025).

A challenge for the SWE approach — The literature to date has not discussed an important challenge to the SWE approach: on essentially every implementation, it implies that

[2] The standard intuition for expecting $\phi < 1$, beyond the fact that population growth has coincided with constant economic growth as noted above, is that when the technological frontier is further advanced, "ideas are harder to find". This at least partially offsets the contribution of past technology to technological development.

[3] Or its "quality", which is modeled as equivalent.

growth could be (and could have been) superexponential, if only we fixed or shrank the range of goods. In fact, under standard assumptions, SWE models tend to produce an "extreme inefficiency result": that **it is feasible to generate infinite utility and/or infinite consumption in arbitrarily little time**, whether the population is fixed or growing.

This point seems to have gone unremarked in part because Proposition 5 of Dinopoulos and Thompson (1998) claims that an optimal growth path exists and can be implemented with an appropriate R&D subsidy; footnote 6 of Aghion and Howitt (1998, ch. 12.2) claims that the "counterintuitive welfare result" that "the optimal number of different products is vanishingly small" can be eliminated with a proposed tweak to the model; and one or the other of these models closely, for our purposes, resembles all subsequent SWE literature. As we will see, if the product range can shrink, the first claim is false and the second is true only given severely diminishing returns to research inputs. Nevertheless, they seem to have closed interest in the efficiency question. No subsequent SWE paper attempts to solve for an optimal growth path.[4,5]

The extreme inefficiency result can be avoided by positing that vertical innovation faces severely diminishing returns to research labor, so that in effect, a larger population can only be employed productively in research if the range of goods widens. Young (1998) takes this approach. In principle, the result can also be avoided by going beyond a traditional lab equipment model—in which research is done by output rather than labor—to one in which the maximum feasible rate of vertical innovation actually *decreases* in population per good. Aghion et al. (2025) make a modeling choice that sometimes has this consequence, on inspection, though it is difficult to see how this implication could be justified.

Outline — In Section 2 I cover research labor models. In Section 3 I cover lab equipment models. I believe that the sections encompass every SWE model with a distinctive feature in any way relevant to the inefficiency results.

Throughout, I aim to show that the tendency for SWE models to exhibit extreme inefficiency is not a mathematical curiosity but follows directly from the central premise of the approach. If growing variety is indeed defusing explosive growth, then if we are not careful we are liable to conclude that slowing or shrinking variety can generate it, and the faster the better. This conclusion is avoidable, but only by making a modeling choice with its own arguably undesirable features.

In Section 4 I briefly evaluate the plausibility of the two conclusions compatible with

[4] Aghion et al. (2025) does solve for optimal policy under constraints.

[5] The point has certainly not gone unremarked simply because it (and perhaps some workaround to it) is too well-known. None of the professors or graduate students at a recent meeting of the Stanford growth seminar had been aware of this point, and it is at the suggestion of one of them that I have written this. The only source of which I am aware that makes a similar point is Davidson (2021), who points out that growth can be hyperbolic in the model of Peretto (2018) with a fixed good range under a parameter restriction.

SWE growth theory: (i) that explosive growth is technologically feasible but goes unrealized due to market failure and (ii) that the returns to innovation diminish more quickly than is usually appreciated. I then note that these possibilities respectively widen the range of answers to how growth will change if production and R&D are much more fully automated, as many technologists forecast they soon will be (Grace et al., 2024). Conclusion (i) implies that the prospect of building agents that can coordinate well enough to avoid the market failure in question—even if they are relatively few and unproductive—constitutes a new channel through which automation could yield explosive growth. Conclusion (ii) implies that the returns to vertical innovation exhibit much more severely diminishing returns than is currently appreciated, such that even full automation may have relatively contained economic impact.

2 Research labor models

2.1 Without severely diminishing returns

This section discusses inefficiency in Dinopoulos and Thompson (1998), Peretto (1998), Peretto and Smulders (2002), and Peretto and Connolly (2007). We will start with a simplification (and in one way a generalization) of Peretto (1998) and discuss its implications in depth. We will then observe that the modifications to this model that make each paper distinctive do not change these implications.

Benchmark: Peretto (1998) (simplified) — A representative individual has log Dixit-Stiglitz preferences:

$$U = \int_0^\infty e^{(\gamma-\rho)t} \ln c_t \, dt, \;\; \rho > \gamma \geq 0; \tag{1}$$

$$c_t = \left(\int_0^{N_t} c_{it}^{\frac{\varepsilon-1}{\varepsilon}} \, di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \;\; \varepsilon > 1, \tag{2}$$

where $c_{it} \geq 0$ denotes the consumption of good i at t, $N_t > 0$ the range of goods available at t, $\rho > 0$ the discount rate, and $\gamma \in [0, \rho)$ the population growth rate.

Production per person of good i is

$$c_{it} = A_{it} L_{it} / L_t, \tag{3}$$

where $L_i \geq 0$ denotes the labor producing good i and $A_i \geq 0$ denotes its productivity.[6]

Let $S \in (0, 1)$ denote the "research share", i.e. the share of the population engaged in research, so that SL_t is the quantity of research labor at t and

$$\int_0^{N_t} L_{it} di = (1 - S) L_t$$

[6]This can be found by rearranging Peretto's equation 5 and letting $A_i \equiv Z_i^\theta$.

is the quantity of production labor. If A_{it} equals a constant value A_t for all $i \in [0, N_t]$, then it is the efficient and the equilibrium outcome for production labor and consumption to be distributed evenly across goods. Fixing S, the consumption aggregate then reduces to

$$c_t \propto A_t N_t^{\sigma}, \quad \sigma \equiv \frac{1}{\varepsilon - 1} > 0. \tag{4}$$

Assume for simplicity that this symmetry holds at time 0.

Process efficiency grows according to

$$\dot{A}_{it} \propto A_{it} P_{it}^{\lambda}, \ \lambda > 0, \tag{5}$$

where $P_i \geq 0$ denotes the quantity of labor doing process research on good i. Since this is the only kind of research we will need to consider, $\int_0^N P_i di = SL$.

Result 1: Infinite utility via double-exponential growth

If we fix N and S, so that P_{it} grows at rate γ for all i, then $\dot{A}_{it}/A_{it}$ itself grows at rate $\gamma\lambda$ for all i, and A_t does as well. More generally, if we fix the fraction of the population engaged in research and allocate it evenly across the product range, symmetry gives us

$$\dot{A}_t \propto A_t (L_t/N_t)^{\lambda}, \tag{6}$$

so that if we enforce a constant $g_N \leq \gamma$, we can sustain a path of $g_{A,t}$ that grows exponentially at rate $\lambda(\gamma - g_N)$. The consumption aggregate c thus grows asymptotically at this double-exponential rate; i.e. $\ln c$ asymptotically grows at rate $\lambda(\gamma - g_N)$. By (1), if

$$g_N < \gamma - \frac{\rho - \gamma}{\lambda}$$
$$\implies \lambda(\gamma - g_N) > \rho - \gamma,$$

the payoff is infinite.

As a corollary, if

$$\rho < \gamma(1 + \lambda), \tag{7}$$

the infinite payoff is achievable even without shrinking N.[7]

[7]Interestingly, condition (7) lies just on the edge of the conventional parameter values $\lambda = 1$, $\rho \approx 0.02$, and $\gamma \approx 0.01$. The utility implications of merely banning new products, in an SWE model, is therefore sensitive to the details. If utility is sublogarithmic in consumption, then even at a lower discount rate, a growth rate that grows exponentially at rate $\lambda\gamma \approx 0.01$ can easily fail to deliver a very large discounted payoff. For intuition, observe that $0.02\, e^{100(0.01)} \approx 0.054$: merely prohibiting new goods in 1925 would, with these parameter values, have raised the growth rate only to 5.4% by today. On the other hand, if we believe that the research population $S_t L_t$ has grown at 4% per year, keeping researchers focused on vertical innovation would have raised the growth rate to over 100% by 2025, and the dramatically faster growth in the meantime would have been highly desirable except under very concave utility functions or high discount rates.

Note that we have not accounted for the fact that foregoing horizontal innovation saves entry costs (and have certainly not assumed that shrinking the range of goods would generate negative entry costs). We will ignore saved entry costs through the rest of the paper as well.

On the model's implications for shrinking N — It is worth emphasizing that if N does shrink, the range of goods shrinks *from above*. Thus if good i exists at time t, then good $j < i$ must as well, as it does on any growth path Peretto considers. That is, the conclusion that it is feasible to produce astronomical growth by discarding products does not involve violating an implicit assumption that simpler goods (or the knowledge used to produce them efficiently) must be maintained to support the production of more complex goods, or anything of that kind. We are simply assuming that if N is small, A_i ($i \leq N$) grows as quickly as if N had always been that small. We are following the letter of the model, and it seems hard to argue that we are not following the spirit.

To elaborate on this point: because A is a *share-weighted* average of the $\{A_i\}$, prohibiting the production of goods $i > \underline{N}$ really does eliminate them in the sense relevant to (4)–(6). A modified model could posit that if a good has ever been developed, it can drag down the productivity of pre-existing goods even after it has been abandoned, but this would be altogether unmotivated. Peretto's justification for including A in the production function for A_i is that "[w]hen one firm generates a new idea to improve its own production process, it also generates general-purpose knowledge that other firms exploit in their own research efforts" (p. 287).

To be sure, if Peretto (1998) is to be well-motivated, the above motivation must be incomplete. Suppose (i) the range of goods rises from $\underline{N}$ to $\overline{N}$, fixing A, and then (ii) A_j doubles only for $j \leq \underline{N}$. Expression (5) implies that $\dot{A}_i$ ($i \leq \underline{N}$) rises by less than if (ii) had occurred alone, even though firm i has access to knowledge produced by firms $j > \underline{N}$. Peretto does not discuss this point; he justifies making $\dot{A}_i$ depend on *average* knowledge across firms (rather than a statistic always non-decreasing in N) only "so that a steady state with constant growth is feasible" (p. 287). However Aghion and Howitt (1998, pp. 407–8) do acknowledge that an SWE model in which one firm's technology has positive spillovers must also allow for negative spillovers in this way, to neutralize scale, and they offer two justifications for accepting this conclusion. First, a wider variety of goods "makes life more complicated". Second, it raises "thin-market transaction costs". The modeling result relevant here—that cutting N allows A_i ($i < N$) to grow as quickly as if N had been low all along—is fully compatible with both arguments.

Result 2: Infinite output in no time

Result 1 relies on Peretto's choice of log utility. If we instead use a utility function featuring more steeply diminishing returns, it is less obvious that exponential growth given (4)–(5) is inefficient at all. A constant, high growth rate may be preferable to one that begins low but grows exponentially.

If we allow N to shrink, however, (4) and (6) imply that for any $T > 0$, there is a feasible growth path with

$$\lim_{t\to T^-} c_t = \infty. \tag{8}$$

For example, suppose the product range shrinks so that

$$N_t \propto (T-t)^{2/\lambda} \quad \text{for} \quad t < T. \tag{9}$$

Since T may be arbitrarily small, assume for simplicity that L is fixed. Substituting $1/N_t$ for P_{it} into (5)—the "$\propto$" allows this if we fix NP_i—and recalling that we will have $A_{it} = A_t$ for all $i \leq N_t$, we have

$$\dot{A}_t \propto A_t(T-t)^{-2} \quad \text{for} \quad t < T.$$

Solving this differential equation for A, and substituting the result along with (9) into (4), yields (8).

The model's claim is not only that infinite output is feasible in finite time but that it is feasible in *arbitrarily little* time, and has been for as long as the model has resembled reality. This is surely implausible, but strictly speaking it does not prove that exponential growth is extremely inefficient for any increasing utility function. This is because, near time 0, sufficiently rapid decreases to N cut consumption more quickly than the not-yet-rapid increases to A raise it, so that c falls temporarily before rising boundlessly near T (and falls nearer to zero the lower T is). Extreme inefficiency would however follow immediately if the model were extended in either of two ways. First, we could allow consumption to be saved and consumed later in even the most marginal way, e.g. by introducing an unproductive form of capital with an arbitrarily high depreciation rate. Second, we could stipulate that the economy can partition itself and instruct one part to carry out the brief research program described above while the other part subsidizes it.

Peretto (1998) proper — The only relevant difference from the "benchmark" model is that, in place of (5), process efficiency grows according to

$$\dot{A}_{it} \propto A_{it}^{1-\psi} A_t^{\psi} P_{it}^{\lambda}, \ \psi \in (0,1), \ \lambda > 0. \tag{10}$$

A_t denotes a consumption-share-weighted, CES/CRS aggregate of A_{it} across $i = 0$ to N_t, so the relationship $A_{it} = A_t$ ($i \leq N_t$) is maintained on the growth paths proposed in Results 1 and 2, and (10) reduces to (5).[8]

Peretto and Smulders (2002) — The only relevant difference from the benchmark is that in place of (5) we have[9]

$$\dot{A}_{it} \propto A_{it}\left(\frac{N_t}{a+N_t}\right)^{\psi} P_{it}^{\lambda}, \ \psi \in (0,1), \ \lambda > 0 \ \text{ for } \ N_t \geq N_0 > -a.$$

[8] Also, incidentally, Peretto considers only the $\lambda = 1$ case.

[9] As above, this can be found by letting $A_i \equiv Z_i^{\theta}$ from the paper's fifth equation on, and the authors consider only the $\lambda = 1$ case. Our a is the authors' $J_0/\delta - N_0$ from their equation 19.

The large parenthetical term may increase or decrease in N_t, depending on the sign of a (though it is asymptotically constant as $N_t \to \infty$). It is microfounded by a model in which firms have positive spillovers on each other's research, but the size of the spillovers depends on how "far apart" the firms are in a space of technology processes, but expansions to the number of firms may also, in a sense, so widen the spread among them that the aggregate spillovers received by any given firm decrease.

Because the magnitude of firm j's spillovers on i does not increase as j grows more technologically advanced (as this would generate scale effects), it does not decrease if we halt j's technological development altogether. We may trivially fix N itself but allocate consumption and research labor over the good range $[0, \underline{N}_t]$ with $\underline{N}_t \to 0$. The inefficiency results follow immediately.

Peretto and Connolly (2007) — The only relevant difference from the benchmark is that in place of (3) we have

$$c_{it} = A_{it}(L_{it} - f)/L_t,$$

where $f > 0$ denotes a fixed cost, in units of labor, that must be paid to produce each good in a given period. Because fewer goods means fewer fixed costs, this tweak only strengthens the inefficiency results.

Dinopoulos and Thompson (1998) — There are two relevant differences from Peretto (1998).

First, $\dot{A}_i$ grows in proportion not to A_i but to a simple (not share-weighted) average of technology levels across goods:

$$A \equiv \int_0^N A_i di \Big/ N. \tag{11}$$

The assumption that it is feasible to shrink the range of goods, as Result 2 and in some cases Result 1 requires, is therefore a substantive one which the paper leaves ambiguous. It depends on whether, if a firm chooses to stop producing good j and/or consumers choose to stop consuming it, j still "constitutes part of the product range", such that stagnation in A_j slows the research process for goods $i < j$ (though it would not have if good j had never been invented). Following the discussion above on the implications of shrinking N, I maintain that there is no coherent motivation—and no motivation proposed in the literature—for the idea that a nonexistent good generates negative spillovers. I will therefore argue that on the most natural reading of the model, shrinking N is feasible.

Second, each A_i does not grow deterministically and continuously as in (5), but in jumps that arrive stochastically with frequency proportional to P_i^{λ}. Because the set of goods is a continuum, however, this stochasticity does not appear in the aggregate, and (6) is precisely maintained.[10]

[10] See the authors' footnote 13 and the notation following their equation 17.

2.2 With severely diminishing returns

This section discusses Young (1998), the only labor-based SWE model to which the inefficiency results do not apply.

The model (simplified) — The model is set in discrete time. The utility function is the discrete-time analog to (1), paired with the standard Dixit-Stiglitz aggregator (2)–(4). Process efficiency however is given by

$$A_{it} - A_{it-1} \propto A_{it-1} \ln(P_{it}/f) \;\text{ for }\; P_{it} \geq f, \tag{12}$$

where f is the fixed labor cost of producing the good. Note that if a good is produced at t all, its process efficiency does not fall.[11]

A continuous-time analog — We will begin by considering a continuous-time analog to the model above. This will allow us to identify how severely the returns to research labor must diminish in order for a research-labor-based SWE model to escape extreme inefficiency, in terms more clearly consistent with the rest of the literature.

Suppose

$$\dot{A}_{it} = mA_{it} \ln(P_{it}/f),\; m > 0. \tag{13}$$

Even the logarithmic research function does not always escape Result 1. The result is now less obvious: infinite output in finite time is now unachievable, and merely enforcing $g_N < \gamma$ now generates linear growth in g_A, so that $\ln c$ grows quadratically rather than exponentially. But if $m > \sigma(\rho - \gamma)$, an infinite payoff is achieved by enforcing $N_t \propto e^{-e^n}$ for $n > \rho - \gamma$. We will work through this point because doing so illustrates why, given log utility, an infinite payoff is guaranteed to be infeasible only if the research function is *sub*logarithmic.

Letting A denote a (perhaps share-weighted) average of $\{A_i\}$, and fixing $P_i = 1/N$,[12]

$$c_t = A_t N_t^{\sigma} = A_0 N_0 e^{\sigma \ln N_t - m \int_0^t \ln N_\tau \, d\tau}.$$

For the exponent to grow at least exponentially at rate $n > \rho - \gamma$, its derivative

$$\sigma \dot{\ln N_t} - m \ln N_t \tag{14}$$

[11] Beyond rearranging the expression of the model, we have simplified it by (i) eliminating a constant (the original would have ...$(\ln(P_{it}/f) - \mu)$ for $P_{it} \geq fe^{\mu}$), (ii) using A_{it-1} in place of Young's "λ_i(max)" construction, and relatedly (iii) removing the possibility of choosing $A_{it} <$ "$\overline{\lambda}(t-1)$". These simplifications have no bearing on how shrinking variety fails to deliver infinite utility or output.

[12] That is, we are normalizing the initial research population to 1 and taking no advantage of the fact that the research population may exhibit sustained growth if $\gamma > 0$.

must also. If $-\ln N$ grows superexponentially, $\dot{\ln N}$ grows more quickly in absolute value and (14) tends to negative infinity. If it grows subexponentially, A is not double-exponential. In the limit, therefore, $\ln N$ must grow exponentially at rate n, so that (14) grows like

$$(\sigma n - m) \ln N_t.$$

This can be made positive with $n > \gamma - \rho$ iff $m > \sigma(\rho - \gamma)$.

Efficiency in the Young model — The Young model proper avoids extreme inefficiency for all parameter values because the discreteness of time amplifies the diminishing returns. Letting g_{it} denote the instantaneous exponential growth rate of process efficiency realized over the course of one period, we have

$$\begin{aligned} e^{g_{it}} &= \frac{A_{it} - A_{it-1}}{A_{it-1}} \propto \ln\left(\frac{P_{it}}{f}\right) \\ \implies g_{it} &\propto \ln\left(\ln\left(\frac{P_{it}}{f}\right)\right). \end{aligned}$$

In the context of the rest of the model, this is enough to guarantee not only that exponential growth is *not extremely* inefficient but that it is not inefficient at all. The optimal growth rate is greater than the equilibrium growth rate—horizontal innovation keeping N proportional to L is supplied at the efficient rate, vertical innovation is undersupplied—but constant (see Young's Section IV).

The intuition for the optimality of $g_N = \gamma$, given sufficiently severe diminishing returns to research labor, is clearest if we imagine the extreme case:

$$\frac{\dot{A}_{it}}{A_{it}} \propto \begin{cases} g, & P_{it} \geq f; \\ 0, & P_{it} < f. \end{cases}$$

3 Lab equipment models

This section discusses inefficiency in the lab equipment SWE literature: Aghion and Howitt (1998, ch. 12.2), Howitt (1999), Howitt (2000), Peretto (2018), and Aghion et al. (2025). In the main, for simplicity, we will not specify preferences and focus only on how the models generate Result 2, trusting that the applications to Result 1 are straightforward.[13]

[13]Incidentally, unlike in the research labor SWE literature, utility in consumption is not universally assumed to be logarithmic, so double-exponential growth does not necessarily have qualitatively extreme welfare implications.

Benchmark: Peretto (2018) (simplified) — Output equals

$$Y_t = \left(\int_0^{N_t} Y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \ \varepsilon > 1; \tag{15}$$

$$Y_{it} = A_{it} L_{it}, \tag{16}$$

so that, given symmetric technology $A_{it} = A_t$ ($i \leq N_t$) and an efficient allocation of labor to each good,

$$Y_t = A_t N_t^{\sigma} L_t, \ \sigma \equiv \frac{1}{\varepsilon - 1} > 0. \tag{17}$$

As before, assume for simplicity that this symmetry holds at time 0. Assume also that the population L grows at rate $\gamma \geq 0$.

Process efficiency for good i grows according to

$$\dot{A}_{it} \propto P_{it} \tag{18}$$

where P_i denotes the rate of investment, in units of *output* rather than labor, into i's process efficiency.[14]

Letting $P_i/(YN) \in (0, 1)$ be fixed and equal across i, so that the research share

$$S \equiv \int_0^N P_i di \Big/ Y$$

is constant and this investment is spread equally across goods, we have

$$\dot{A}_t \propto Y_t/N_t \propto A_t N_t^{\sigma-1} L_t. \tag{19}$$

If $\sigma < 1$, then with $\lambda \equiv 1 - \sigma$, Result 2 applies immediately.

If $\sigma \geq 1$—i.e. $\varepsilon \leq 2$—vertical innovation does not accelerate when variety shrinks. In this case, when variety doubles, output weakly more than doubles, so even lab equipment *per good* weakly increases. Variety is unambiguously desirable. Precisely for this reason, however, the SWE mechanism for defusing superexponential growth fails. Given $\gamma > 0$, output per person grows double-exponentially, not only on a feasible path attainable by fixing N but in equilibrium. For this reason, Peretto restricts his attention to the $\sigma < 1$ case.[15]

Note from (18) that the rate of vertical innovation is assumed to be linear in investment. That is, in the notation of the previous section, it is assumed that $\lambda = 1$. This is because here, technology supports technological development only via the channel that $Y \propto A$ and $\dot{A}_i \propto P_i \propto Y$. If the function from P_i to A were not linear, the model would not

[14] We are using A to denote Peretto's "Z^{κ}", but here considering only the $\kappa = 1$ case. The $\kappa > 1$ case is discussed below.

[15] See p. 54, especially the equivalence to the Romer model given $\sigma \geq 1$ discussed in footnote 6.

satisfy the central desideratum that, fixing L and N, output grows exponentially through vertical innovation.

Peretto (2018) proper — The only relevant difference between Peretto (2018) proper and the simplification above is that, in place of (18), he uses[16]

$$\dot{A}_{it} \propto A_{it}^{\zeta} P_{it}, \ \zeta \geq 0,$$

so that, fixing $P_i/(YN) > 0$,

$$\dot{A}_t \propto A_t^{1+\zeta} N_t^{\sigma-1} L_t.$$

The $\zeta > 0$ case yields a result more extreme than in any model yet: even fixing rather than shrinking N, process efficiency and output can grow hyperbolically.

With $\zeta > 0$, the desideratum that exponential growth is feasible fixing L and N is compatible with exponentiating P_i by $\lambda \geq 1 - \zeta$. Hyperbolic growth with fixed variety remains feasible as long as the inequality is strict, and Result 2 is maintained in any case.

Howitt (1999) — Output equals

$$Y_t = \int_0^{N_t} Y_{it}^{1-\sigma} di, \ \sigma \in (0,1);$$
$$Y_{it} = A_{it}^{\frac{1}{1-\sigma}} L_{it}.$$

Given symmetric technology $A_i = A$ ($i \leq N$) and an efficient allocation of labor $L_i = L/N$ ($i \leq N$), this would yield

$$Y_t = A_t N_t^{\sigma} L_t^{1-\sigma}.$$

Implicitly, output exhibits constant returns to scale in (i) the Dixit-Stiglitz aggregate (15) of intermediate goods i and (ii) a fixed factor whose share coincides with σ. Note that A_i no longer precisely denotes i's process efficiency, but that by construction output remains linear in the (weighted) average A of the $\{A_i\}$.[17]

Howitt assumes that goods' process efficiencies grow stochastically, so the $\{A_i\}$ are not equal. As in Dinopoulos and Thompson (1998), however, the process is such that, given equal research investment per product, the shape of the distribution is time-invariant.[18] The efficient allocation maintains

$$\int_0^N Y_i^{1-\sigma} di \ \propto \ N \cdot \left(A^{\frac{1}{1-\sigma}} L/N\right)^{1-\sigma}$$

[16] After the change of variables noted in footnote 14 above, our ζ is Peretto's $1 - 1/\kappa$.

[17] For consistency with the other sections, our σ is Howitt's $1 - \alpha$. Also, Howitt defines A_i so that Y_i is linear in A_i but Y is nonlinear in average process efficiency; we use a change of variables to do the reverse.

[18] Asymptotically or, if the initial distribution is the asymptotic distribution, precisely.

in the efficient labor allocation, and thus

$$Y_t \propto A_t N_t^{\sigma} L_t^{1-\sigma}.$$

As in Peretto (2018), process efficiency for good i grows essentially according to (18). The only difference is that $\dot{A}_i$ is linear not in P_i but in $P_i \cdot A_i / \max_i(\{A_i\})$, but by invariance the numerator and the denominator grow at the same rate over the long run. The aggregate result, fixing the research share S, is a deterministic process familiar from (19):

$$\dot{A}_t \propto Y_t/N_t \propto A_t N_t^{\sigma-1} L_t^{1-\sigma}.$$

Because $\sigma < 1$, Result 2 follows immediately.

Aghion and Howitt (1998, ch. 12.2) and Howitt (2000) — Output equals[19]

$$\begin{aligned} Y_t &= N_t^{\alpha-1}\Big(\int_0^{N_t} Y_{it}^{\alpha}\, di\Big) L_t^{1-\alpha} \cdot \Big(\frac{N_t}{L_t}\Big)^{\sigma}, \ \alpha \in (0,1),\ \sigma \in [0,1); \qquad (20) \\ Y_{it} &= A_{it}^{1/\alpha} K_{it}, \end{aligned}$$

where the capital stock $K = \int_0^N K_i di$ grows with saved output in the usual way, but this will not be relevant for our purposes.

As in Howitt (1999), process innovations arrive stochastically, but in a way that maintains an invariant distribution of $\{A_i\}$ and, in the efficient capital allocation,

$$\begin{aligned} \int_0^N Y_i^{\alpha} di \ &\propto\ N \cdot \Big(A^{1/\alpha} K/N\Big)^{\alpha} = N^{1-\alpha} A K^{\alpha} \\ \implies\ Y_t &\propto A_t K_t^{\alpha} L_t^{1-\alpha} \cdot (N_t/L_t)^{\sigma} \end{aligned}$$

where A as usual may be an average or weighted average of the $\{A_i\}$.

Remarkably, in the model studied throughout the body of the textbook chapter and the entirety of Howitt (2000), the $(N/L)^{\sigma}$ term at the right of (20) is absent; σ equals zero. The $N_t^{\alpha-1}$ term at the left precisely nullifies the benefits of variety that the Dixit-Stiglitz-style aggregate is usually designed to capture, and inventing new goods effectively just fragments "A" into smaller pieces. In footnote 6 of the chapter, the authors acknowledge that "[l]iterally, the model implies that the optimal number of different products is vanishingly small", but say that this "counterintuitive welfare result" can be eliminated with the tweak above.[20]

[19]For consistency with the other sections, we have used σ for the authors' β and let Y_i stand for the authors' $A_i^{1/\alpha} x_i$. We have then redefined A_i for the reasons in footnote 17 above.

[20]They argue that placing an L in the denominator, so that doubling population and variety has no impact on output per person, "might be justified by the fact that the variety of different tastes... expands as people become more numerous" or by "thin-market costs". Removing this penalty to population (or making it weaker than σ) would affect the equilibrium, but it has no bearing on the inefficiency results.

In the case of Howitt (2000), this is always false. Given a constant research share S and an equal allocation of process efficiency investment across goods, A evolves according to

$$\dot{A}_t \propto Y_t/N_t$$

for the usual reasons. By assumption $\sigma < 1$, so Result 2 follows at once.[21]

In the chapter itself, A evolves according to

$$\begin{aligned}\dot{A}_t &\propto A_t R\Big(\frac{SY_t}{A_t N_t}\Big) \text{ with } R(0)=0, R'>0, R''<0 \\ &= A_t R\big(SK_t^{\alpha} L_t^{1-\alpha-\sigma} N_t^{\sigma-1}\big).\end{aligned}$$

If R' diminishes no more steeply than a power function, Result 2 still follows, as seen in the discussion of Peretto (2018) proper. More generally, precisely as in the continuous-time analog to Young (1998), only if $R(\cdot)$ is sublogarithmic, or logarithmic under certain parameter values.

Aghion et al. (2025) — Output equals (15)–(17), as in Peretto (2018). Process efficiency for good i grows according to[22]

$$\dot{A}_{it} \propto A_{it}\Big(P_{it} A_t^{\frac{1-\sigma}{\sigma}} A_{it}^{-\frac{1}{\sigma}} L_t^{-\sigma}\Big)^{\lambda},\ \lambda > 0, \tag{21}$$

where A is a CRS, CES aggregate of $\{A_i\}$ with elasticity of substitution greater than 1.[23] As usual, assume symmetry, replace A_i with A, and, fixing S, substitute Y/N for P_i to get

$$\dot{A}_t \propto A_t\big(L_t/N_t\big)^{\lambda(1-\sigma)}.$$

As with Peretto (2018), Result 2 holds iff the exponent on N is negative, which it is iff $\sigma < 1$ ($\varepsilon > 2$). Now, however, if $\sigma > 1$, a growing population does not drive double-exponential growth if $g_N < \gamma$, because the exponent on L is negative. The model avoids extreme inefficiency in the $\sigma \geq 1$ ($\varepsilon \in (1,2]$) case, but only by constructing the research function precisely such that, under that condition, *the maximum feasible rate of vertical innovation decreases in population per product.*[24]

[21] In fact, the paper assumes for simplicity that utility is linear in consumption, so even a cut to the product range big enough to raise the growth rate above the discount rate generates infinite utility.

[22] This follows from rearranging the authors' equation 6, using ε for their θ (so our σ equals their $1/(\theta-1)$) and λ for their $1/(1+\zeta)$. They focus only on the $\lambda < 1$ ($\zeta > 0$) case.

[23] As with Dinopoulos and Thompson (1998), we posit that N may shrink, so that if research and production are restricted to goods up to N_t, A_t equals $\{A_{it}\}$ ($i < N_t$). Here, however, the paper is explicit that exit is possible and that discontinued goods are removed from the technology aggregate.

[24] Another distinctive feature of the paper is that it posits that a stream of investment is needed to keep a given range of goods "alive", somewhat analogous to the fixed costs of Peretto and Connolly (2007) or Young (1998). This is the motivation for the paper: the fixed costs are interpreted as the costs of developing new goods once we run out of further process efficiencies for the old goods. These costs do not change the conditions under which Result 2 obtains, but because a smaller good range requires smaller fixed costs, they only strengthen the result.

In short, extreme inefficiency may be avoided (under some parameter values) if we go beyond a standard lab equipment model—in which output is the only input to research and labor offers no direct contribution—and posit that (under said parameter values) the existence of human beings generates a direct negative contribution to the inputs to vertical innovation, which can be relieved only by creating variety. Here, this direct cost has elasticity σ, as can be seen from (21). If $\sigma > 1$, this negative contribution grows more quickly in L/N than the indirect benefit that a larger population creates more output to devote to research. The authors justify this modeling choice as "captur[ing] the effect that a larger market leads to faster economic growth", but it is unclear how it could capture this effect, since it is in fact a stipulation that larger markets make growth more costly. It is also unclear why we should expect the elasticity of this cost to be precisely σ, so that population growth puts a net drag on technological development precisely when it would otherwise suffice to render growth double-exponential.

4 Discussion

We have seen that the SWE approach to growth theory implies either (i) that vertical innovation faces severely diminishing returns or (ii) that explosive growth is feasible but goes unrealized due to a colossal market and/or policy failure. This "dilemma" has implications for the plausibility of the SWE approach. If we accept that the approach is broadly accurate nonetheless, the dilemma in turn has significant implications for the growth impacts of automation.

4.1 Implications for second-wave endogenous growth theory

Plausibility of severely diminishing returns — The SWE literature since Young (1998) has entirely abandoned the assumption of sublogarithmic returns to investments in process efficiency.[25] Indeed, even Young does not assume sublogarithmic returns explicitly. The most common assumption is that the returns are linear. This is presumably because such steeply diminishing returns are considered intuitively implausible and because firm-level studies of the elasticity of productivity increases to research investment do not typically find them.[26]

That said, it does not seem unreasonable to suppose that the elasticity quickly approaches zero outside the observed range. No research team, however large, could double process efficiency in five minutes.

Plausibility of extreme inefficiency — The alternative possibility is that "explosive growth" along the lines of Result 2 is and has long been technologically feasible. This possibility

[25] With the partial exception of Aghion and Howitt (1998, ch. 12.2), as noted.

[26] See e.g. Hall and Mairesse (1995), Lanjouw and Schankerman (2004), Klette and Kortum (2004), and Akcigit and Kerr (2018).

may be highly counterintuitive, but if it is true it is among the most important facts of all time. It is worth evaluating seriously.[27] Market and policy failures can destroy markets even in the developed world, and have kept growth rates low in some developing countries for decades even after many examples of rapid catch-up growth have offered case studies for others to follow. Ultimately, however, this possibility appears untenable for at least three reasons.

First, if explosive growth is feasible, the size and speed of its benefits drastically exceed those of resolving any other case of stagnation or market failure. A country may stagnate if a small increase to its growth rate would require investments or policy changes with large up-front costs, especially if the costs accrue to its current elite (see e.g. Acemoglu and Robinson (2012)), but any such costs will pay for themselves if the growth impact is sufficiently large and rapid. Moreover, the history of Mao's China demonstrates that it is feasible for a modern state to enforce severe "negative horizontal innovation", restricting output to a narrow range of goods. History also of course offers many examples of large, successful, publicly sponsored research projects in narrow domains, such as the Manhattan Project, the Apollo Program, or the Human Genome Project. It is hard to see why no state would be able to afford whatever costs are associated with sustaining negative horizontal innovation and a large research program at the same time.

Second, the logic of an SWE model implies (absent sublogarithmic returns to investments in process efficiency) that, even if a single firm or community refused to trade with the rest of the world, it would be able to generate explosive growth on its own. This putatively does not happen only because the members of any such group would face a constant temptation to depart and establish monopoly over a new product, whose rents would be astronomical due to its complementarity to the few but plentiful products produced by the rest. The solutions to this kind of coordination problem are straightforward: e.g. a consortium agreement ensuring that, if any party withdraws, the project is canceled. It is hard to imagine that no group in the world, over generations, could have implemented one.

Third, many people already choose careers in part on the basis of their social impact. Many more would accept below-market pay to join a team—indeed, surely many would pay to join a team—that could open the gates of heaven.

Alternative preferences — If we reject both the possibilities above, we reject the SWE literature, at least in its current form. The most common alternative, as noted in the introduction, is semi-endogenous growth theory. Semi-endogenous growth models maintain the Dixit-Stiglitz preference specification, but propose an alternative research function which implies that economic growth can be sustained only by population growth.

Another possibility is that the SWE-style research function (on both dimensions) is

[27] Indeed, one of the primary motivations for the SWE approach to growth theory is that it is said to match the evidence on how policy can affect the growth rate: see the introduction of Aghion et al. (2025) for an up-to-date statement of the case. If this is true, its implication that a new policy paradigm could send growth to infinity is worth taking especially seriously.

roughly accurate, but that the Dixit-Stiglitz aggregator leads us badly astray. In particular, if marginal utility diminishes rapidly enough in the consumption of each good, consuming a wider range of goods is sometimes preferable even to arbitrarily large quantities of a narrower range of goods. Then even if explosive growth in a vanishing range of goods is feasible, as in Result 2, it may be preferable to widen the range of goods instead.[28] In combination with fixed labor costs to maintaining a given product line, it may be optimal to keep variety proportional to population, and for stagnant populations in the long run to engage only in vertical innovation.[29]

Because the required preferences are highly nonhomothetic, however, they do not admit a consumption aggregate at all. There is then no straightforward sense in which economic growth has historically been exponential, the central stylized fact which SWE and semi-endogenous models both seek to explain. Though an alternative preference specification may salvage parts of the SWE approach, therefore, it should arguably motivate a reexamination of growth theory more fully.

4.2 Implications for growth after automation

Severely diminishing returns — In a semi-endogenous model, economic growth proceeds at roughly its maximum possible rate. Growth is, in a technological sense, more difficult than in an SWE model: it is constrained not by equilibrium decisions to expand variety but simply by a lack of people. Precisely for this reason, however, a semi-endogenous model predicts that in the event of full automation—where output can be turned into robots capable of performing every task in research and production—growth will accelerate dramatically. In fact, even sufficient partial automation renders growth hyperbolic.[30]

Now consider a simple SWE-style model, under full automation, in which vertical innovation faces severely diminishing returns. Consumption good i is produced by the "bots" B_i allocated to producing it, which they do with efficiency A_i:

$$c_{it} = A_{it} B_{it},$$

so that under symmetry the Dixit-Stiglitz aggregator equals

$$c_t = A_t N_t^{\sigma} B_t,$$

where $\int_0^N B_i di = B$ and as usual $\{A_i\} = A$. Fixing the share of robots engaged in vertical

[28] Work in progress with Chad Jones explores this and related points. For a very early draft containing some of the intuition, see Trammell (2024).

[29] I thank Pete Klenow for making this point.

[30] See Aghion et al. (2019), Section 4.1, example 3. For a corrected proof see Trammell and Korinek (2023), Section 6.

innovation, process efficiency for good i grows according to[31]

$$\dot{A}_{it} = A_{it} R\big(B_t / N_t\big) \;\; \text{with} \;\; R' > 0, \lim_{x \to \infty} R(x) = \bar{g}.$$

Like consumption goods, bots are themselves a produced good. Fixing the shares of bots engaged in bot production and bot vertical innovation, the bot population grows according to

$$\dot{B}_t \propto A_{\text{BOT},t} B_t,$$
$$\dot{A}_{\text{BOT},t} = A_{\text{BOT},t} R(B_t).$$

We ignore depreciation because, with B growing superexponentially, the exponential depreciation process is irrelevant in the limit.

In the limit, A grows at a rate no greater than $\bar{g}$ and B grows double-exponentially at a rate proportional to $e^{\bar{g}t}$. The final source of consumption growth is then g_N.

Fix the share of bots engaged in developing new goods. Suppose that new goods inherit average process efficiency, as roughly in the existing SWE literature (so that the technology aggregate A grows like A_i for each i in production, as stipulated). If, by analogy to one strand of the literature,

$$\dot{N}_t \propto B_t,$$

then N grows like an exponential integral—slower than double-exponentially. The assumption that "new goods per bot" is constant might be justified on the grounds that developing a new good j and a process for it that is as efficient as the existing A grows more difficult as A rises, in a way that just offsets whatever positive spillovers come from the technology associated with high process efficiency in existing goods. Alternatively, following another strand, we might simply specify

$$N_t \propto B_t.$$

Then N grows double-exponentially with B. In either case, c grows at most double-exponentially like $e^{e^{\bar{g}t}}$.

In this framework, we know that $\bar{g}$ exceeds the current rate of process efficiency growth (say 1% per year, if the economic growth rate is roughly split between vertical and horizontal innovations). We also know that $R'(x)$ begins to diminish rapidly for x not too far above $R^{-1}(1\%)$, or else a result qualitatively similar to Result 2 would obtain; the burst of growth achievable by shrinking variety would be finite but dramatic. It is difficult to say much more. In any case, double-exponential growth is eventually arbitrarily rapid—we are imposing no limits on the rate at which robots will ultimately be able to self-replicate or build the things we value—but there is a big difference between a vertical asymptote and a growth rate that itself grows at, say, 10% per year.

[31]Positing that $R(\cdot)$ is upper-bounded simplifies the analysis. Unbounded but sufficiently concave research functions can likewise yield the conclusion below that, on an efficient path, output grows double-exponentially instead of exhibiting a vertical asymptote.

Note how the central mechanism behind the result, an extremely concave research function, would be incompatible with the semi-endogenous framework (without further modifications). For exponentially growing research labor to drive exponential growth in $\dot{A}$, $R(\cdot)$ can be at worst power-functional.

Extreme inefficiency — Briefly, suppose we accept the SWE framework but reject the conclusion that vertical innovation faces severely diminishing returns, maintaining instead that growth has historically been extremely inefficient because of the coordination failures discussed in the previous subsection. Then automating process efficiency research, even for a narrow band of goods, yields Result 2 as long as the automated researchers can be instructed to coordinate. Hyperbolic growth does not require the automation of any part of production; the "bots" do not have to be self-replicating, numerous, or able to produce any of the final goods we value.

References

Acemoglu, Daron and James A. Robinson, *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York: Crown, 2012.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–51.

— **and** — , *Endogenous Growth Theory*, Cambridge, MA: MIT Press, 1998.

— **, Antonin Bergeaud, Timo Boppart, and Jean-Félix Brouillette**, "Resetting the Innovation Clock: Endogenous Growth through Technological Turnover," 2025. Working paper.

— **, Benjamin F. Jones, and Charles I. Jones**, "Artificial Intelligence and Economic Growth," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, 2019, pp. 237–282.

Akcigit, Ufuk and William R. Kerr, "Growth through Heterogeneous Innovations," *Journal of Political Economy*, 2018, *126* (4), 1374–1443.

Davidson, Tom, "Report on Whether AI Could Drive Explosive Economic Growth," Technical Report, Open Philanthropy 2021.

Dinopoulos, Elias and Peter Thompson, "Schumpeterian Growth without Scale Effects," *Journal of Economic Growth*, 1998, *3* (4), 313–35.

Grace, Katja, Harlan Stewart, Julia Fabienne Sandkühler, Stephen Thomas, Ben Weinstein-Raun, and Jan Brauner, "Thousands of Authors on the Future of AI," 2024. Preprint.

Grossman, Gene and Elhanan Helpman, *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press, 1991.

Hall, Bronwyn H. and Jacques Mairesse, "Exploring the Relationship between R&D and Productivity in French Manufacturing Firms," *Journal of Econometrics*, 1995, *65* (1), 263–93.

Howitt, Peter, "Steady Endogenous Growth with Population and R & D Inputs Growing," *Journal of Political Economy*, 1999, *107* (4), 715–30.

_ , "Endogenous Growth and Cross-Country Income Differences," *American Economic Review*, 2000, *90* (4), 829–46.

Jones, Charles I., "R&D-based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 495–525.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, *65* (6), 1389–1419.

Lanjouw, Jean O. and Mark Schankerman, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *The Economic Journal*, 2004, *114*, 441–65.

Peretto, Pietro, "Technological Change and Population Growth," *Journal of Economic Growth*, 1998, *3* (4), 283–311.

_ **and Michelle Connolly**, "The Manhattan Metaphor," *Journal of Economic Growth*, 2007, *12* (4), 329–350.

_ **and Sjak Smulders**, "Technological Distance, Growth and Scale Effects," *The Economic Journal*, 2002, *112*, 603–624.

Peretto, Pietro F., "Robust Endogenous Growth," *European Economic Review*, 2018, *108*, 49–77.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98* (5), 71–102.

Segerstrom, Paul, "Endogenous Growth without Scale Effects," *American Economic Review*, 1998, *88* (5), 1290–1310.

Trammell, Philip, "New Products and Long-term Welfare," 2024. In progress.

_ **and Anton Korinek**, "Economic Growth Under Transformative AI," 2023. NBER Working Paper No. 31815.

Young, Alwyn, "Growth without Scale Effects," *Journal of Political Economy*, 1998, *106* (1), 41–63.