The Agentic Production Economy: Intelligence, Capital, and Autonomous Enterprise

Paper II of *Foundations of Agentic Economics*

Matthew Long
The YonedaAI Collaboration
YonedaAI Research Collective
Chicago, IL
matthew@yonedaai.com

August 2026

Abstract

Paper I of this series treats the obligation as the primitive economic object and money as the residual left over after clearing, $M_t = \mathcal{R}(\mathcal{E}_t)$. This paper asks what changes when some of the nodes in that obligation graph are artificial agents that produce, contract, and settle on their own account. Our answer is deliberately narrow, because the broad version of the claim is wrong. The novelty is not in the production function: an artificial intelligence that merely produces is capital, and the existing literature on automation already describes it well. The novelty is in the location of the contracting node. A machine that is a factor input cannot issue an obligation; an agent whose forward output can be underwritten can. We therefore develop the economics of the second case. We take the series' production decomposition $\mathcal{P} = H + A + K + E + I$ and separate its two readings, showing that as a statement about technology the additive form is the perfect-substitutes upper envelope of the constant-elasticity family and so overstates achievable output whenever inputs are complements. We give economic capacity $C_i(t)$ an operational definition as the largest obligation stock an agent can credibly carry, argue that it replaces the budget constraint as the binding restriction on agent behaviour, and show that money holdings are the special case in which underwriting is performed by a third party instead of by the counterparty. We then develop the Agentic Enterprise as an organisational form and argue, against the natural Coasean reading, that cheap machine negotiation does not dissolve the firm: it relocates the firm's boundary from communication cost to verification and underwriting cost, and it migrates agency costs from monitoring to specification. Four objections are treated at length instead of dismissed: reliability failure that does not diversify away because agents share model lineage, the Minskyan instability of lending against projected machine output, the absence of legal standing for an agent balance sheet, and the concentration risk created when a small number of model providers hold what amounts to an undisclosed senior claim on every dependent agent's production. The paper is a proposal for how to think about a system that does not yet exist in the form described, not a report of settled results.

1 Introduction

Economics has a long habit of absorbing new machines into the capital stock and moving on. The spinning frame, the dynamo, the mainframe and the industrial robot all entered the theory as

K, and the interesting questions promptly became questions about factor shares, substitution elasticities and the pace of adoption. The habit has served the discipline well. There is a serious case, developed carefully by Zeira (1998) and by Acemoglu and Restrepo (2018, 2019), that artificial intelligence should be absorbed in exactly the same way: as a technology that converts tasks previously performed by labour into tasks performed by capital, with consequences for wages, employment and factor shares that existing models already describe. Aghion et al. (2019) show how far one can get by treating AI as an input to the production of ideas, and Nordhaus (2021) shows how demanding the conditions for a growth singularity actually are once one takes the substitution structure seriously.

This paper takes that case seriously and then argues that it is incomplete in one specific respect. What is new about an artificial agent is not that it produces. Machines have produced for two centuries. What is new is that an artificial agent can hold the contract.

The distinction matters because a factor input and a contracting party occupy different positions in economic theory. A lathe cannot promise. It cannot be the obligor on an edge of the obligation graph developed in Paper I of this series. When a lathe is financed, the promise runs from the lathe's owner to the lender, and the lathe appears only as collateral. An artificial agent that discovers demand, negotiates terms, procures its own compute, delivers verified output and settles the resulting obligations occupies the obligor position directly. The economics that changes is therefore contract theory and monetary economics, not production theory. That is a smaller claim than the one usually made about AI and the economy, and we think it is a more defensible one.

It is also a conditional claim, and the condition should be stated at the outset, not discovered in Section 8. If an artificial agent can never be the obligor in economic substance, so that every promise it appears to make is legally and practically a promise by its operator, then the obligation graph has not changed shape at all. The operator's node is still the node, the agent is a tool the operator uses to draft, negotiate and produce, and automating those functions is capital deepening in managerial and legal work. The framework in this paper would then reduce to a vocabulary for describing automation, and the automation literature would describe it better. Section 7.2 argues that this is the situation today in every jurisdiction we are aware of. Everything that follows should be read as conditional on that situation changing, and Section 8 sets out the two routes by which it might.

Paper I establishes the framework this paper builds on. The economy is a time-dependent directed graph $\mathcal{E}_t = (V, E_t)$ whose vertices are humans, firms, governments, machines and artificial agents, and whose edges $e_{ij} = (q, u, t, c, r)$ are obligations characterised by quantity, unit of account, maturity, conditions and risk. Every financial asset is another party's liability, so the closed system sums to zero. Money is not a primitive but a graded property of claims, and the quantity of it required at any moment is the residue left after permissible clearing and netting, $M_t = \mathcal{R}(\mathcal{E}_t)$. That framework is agnostic about who occupies a vertex. This paper supplies what Paper I leaves open: the conditions under which a counterparty accepts an edge from an artificial agent at all. Capacity is what makes an agent's edges credible, and the greater part of what follows is an attempt to say what capacity is and how it behaves.

What the paper argues

The argument proceeds in five moves.

First, we take the series' production decomposition $\mathcal{P} = H + A + K + E + I$ and insist on separating two things that the notation runs together. Read as an accounting identity over the value of services delivered in a period, the additive form is unobjectionable and carries no economic content beyond bookkeeping. Read as a statement about technology, it asserts that human capability, autonomous intelligence, physical capital, energy and information are perfect substitutes at the margin. We show in Proposition 1 that the additive form is the upper envelope of the constant-elasticity-of-substitution family, so that for any degree of complementarity actual output is strictly lower. The additive notation is not neutral, and a paper that uses it without saying so is quietly assuming the most optimistic case.

Second, we treat autonomous intelligence as produced capital instead of as an exogenous factor. The term A is manufactured out of compute capital and energy, which makes the production structure recursive and puts a physical floor under the whole system.

Third, we give economic capacity an operational definition. In the series brief, $C_i(t)$ is a function of capability, reliability, capital, energy, reputation and expected future production. We define it as the largest obligation stock the agent can carry while keeping its probability of default below a threshold, which turns out to equal a haircut applied to discounted expected future production. Capacity then replaces the budget constraint as the restriction that binds, and cash balances become a special case of capacity in which the underwriting is done by a third party.

Fourth, we develop the Agentic Enterprise as an organisational form and ask what happens to the theory of the firm. The naive Coasean prediction is that near-zero negotiation cost dissolves firm boundaries into atomistic contracting. We argue this is wrong, for a reason internal to the Coasean tradition itself: what cheap machine negotiation reduces is search and drafting cost, not verification cost, and Grossman and Hart (1986) and Hart and Moore (1990) located the firm's rationale in control over non-contractible decisions and not in the price of writing contracts. Agentic contracting should therefore push activity into markets precisely where output is cheaply verifiable and consolidate it where it is not.

Fifth, we take four objections seriously enough to develop them formally. Reliability failure does not diversify away when agents share model lineage, which we make precise in Remark 11. Credit extended against projected machine output has a Minskyan structure, and we exhibit a simple recursion in Proposition 10 whose stability depends on the elasticity of demand for agent output in an uncomfortable way. An agent balance sheet has no legal standing, and the workarounds available under current entity law are less reassuring than they first appear. Finally, when a small number of providers supply the intelligence that every agent's capacity depends on, those providers hold something functionally equivalent to a senior claim on the entire dependent production chain, and it appears on nobody's books.

Relation to the literature

The economics of artificial agents as economic actors, as distinct from AI as a production technology, has begun to develop only recently. Hadfield and Koh (2025) give the closest treatment to the position taken here, framing artificial agents as parties that plan, negotiate and operate under institutional constraints, and asking what institutions such agents require. Tomasev et al. (2025) propose sandboxed agent economies as a design response to systemic risk from agent-to-agent transaction layers. Aldasoro and Desai (2025) provide something rarer, an empirical study of language-model agents managing liquidity buffers and settlement timing in a payment system, which is direct evidence that agents can manage balance-sheet positions and not merely produce outputs. Yang et al. (2025) develop an auction-based marketplace framing.

We should be honest about the limits of this coverage. We found no central-bank or multilateral research modelling autonomous artificial firms as such, as opposed to AI deployed inside existing financial institutions or AI as a labour-market shock. The institutional literature on agentic production is thin, and this paper does not have a body of established results to lean on. Where we assert something about how such a system would behave, we are proposing rather than reporting.

The rest of the paper is organised as follows. Section 2 places production inside the obligation graph and fixes notation. Section 3 treats the production decomposition. Section 4 develops economic capacity. Section 5 treats credit against future production and its instability. Section 6 develops the Agentic Enterprise and the theory of the firm. Section 7 treats the evolution of the economic unit and the legal standing problem. Section 8 consolidates objections. Section 9 states what would falsify the framework. Section 10 connects to Papers III through V.

2 Production Inside the Obligation Graph

2.1 What Paper I supplies

Paper I's central move is to deny money the status of an economic primitive. The primitive is the obligation relation, and money is a grade of claim and not a category apart from credit. The grading is captured by a money-quality function

$$\mathcal{M}(x) = f(\text{acceptability}, \text{liquidity}, \text{credit quality}, \text{transferability}, \text{settlement finality}, \text{stability}), \quad (1)$$

under which a central bank liability scores near the top, a bank deposit slightly lower, a corporate trade credit lower again, and a personal promissory note near the bottom. Nothing in the framework requires a discontinuity anywhere along that scale, and Paper I argues that no such discontinuity exists. The intellectual lineage runs through Innes (1913), who argued that credit and credit alone is money and that trade generates mutual obligations of which only the residue is settled; through Knapp (1924), who located the acceptability of state money in the state's willingness to receive it; through Graeber (2011), whose anthropological record shows webs of obligation with periodic reckoning preceding coined money; and through Mehrling (2011), whose hierarchy of money makes each tier's money the next tier's credit, unified only at settlement.

The conceptual sequence Paper I develops is

$$\text{Production} \to \text{Claim} \to \text{Credit} \to \text{Money} \to \text{Settlement}. \quad (2)$$

Production comes first in that sequence, and Paper I says almost nothing about it. That is this paper's assignment.

2.2 What this paper adds

Consider an edge e_{ij} in E_t under which agent i owes agent j a quantity q at maturity t. Paper I takes the existence of the edge as given and studies what the graph does. But an edge exists only because j agreed to accept i's promise instead of demanding immediate settlement. That acceptance decision is an economic problem in its own right, and in the source formulation it is written

$$\mathrm{P}(\text{accept } O_{ij}) = g\big(\text{solvency}_i,\ \text{reputation}_i,\ \text{expected output}_i,\ \text{collateral}_i,\ \text{network guarantees}\big). \quad (3)$$

Four of the five arguments to g are statements about i's productive future, not about i's current holdings. This is the sense in which production underwrites the obligation graph.

This raises a question that should be answered before it becomes an objection. Paper I already establishes a condition governing which edges can exist, showing that the feasible edge set shrinks as the cost of verifying an obligation's conditions rises relative to the surplus the obligation generates. If Paper I tells us which edges are feasible, why does this paper need a separate condition for whether one is accepted? Because the two constraints are different and both bind. Feasibility is a property of the obligation: can this promise be written and enforced at a cost the transaction can bear? Acceptance is a property of the obligor: will this particular counterparty take this particular promise from this particular issuer? An obligation can be perfectly cheap to verify and still be refused because the issuer cannot support it, and an issuer of unimpeachable capacity can find an obligation infeasible because no one can check whether its conditions were met. Paper I supplies the first constraint and this paper supplies the second. Section 5.5 shows what happens when they interact, since the same verification cost that governs feasibility in Paper I also governs whether forward production can be attested well enough to establish capacity here. A graph of claims on future production is only as sound as the production is credible, and the credibility of machine production is the subject matter of this paper.

The five-layer architecture Paper I proposes makes the dependency explicit. The capability layer records what each actor can produce; the obligation layer records promises created in exchange for production; the trust layer holds probabilistic estimates that obligations will be honoured; the clearing layer nets reciprocal obligations continuously; and the settlement layer handles the residue. The capability layer sits underneath everything else. If it is mis-measured, every layer above it inherits the error, and the clearing layer will net obligations that were never sound to begin with.

2.3 Notation

We keep the series' notation exactly, including one overload that we flag without fixing. The symbol E carries three meanings across the series: the edge set E_t of the obligation graph, energy as a productive input, and economic power E_i. The series brief and knowledge base both use all three. Renaming would break consistency with the sibling papers, so we retain them and disambiguate by context and by subscript. Paper I adopts the same convention and carries the same warning, so a reader moving between the two is told twice.

The letter C carries a second collision that is internal to this paper and that we should dispose of before it causes trouble. We write $C_i(t)$ for economic capacity throughout, in the sense of Definition 4. The agent balance sheet inherited from the source material, equation (14), uses C_i in its third slot for productive capabilities, which is a different object: capabilities are what the agent can do, while capacity is the obligation stock those capabilities will support once discounted and haircut. Where the distinction matters we write $C_i(t)$ for capacity and refer to the balance-sheet entry as the capability component. Paper III uses C_i for final consumption in its per-agent tax form, and Papers I and III both use L_i where we use it for liabilities and Paper III uses it for land value. None of these are the same variable, and each paper states its own scope. Table 1 collects the symbols used here.

3 The Production Decomposition

3.1 Two readings of an additive form

The series brief replaces the firm-centric production model with

$$\mathcal{P} = H + A + K + E + I, \tag{4}$$

where H is human capability, A autonomous intelligence, K physical capital, E energy and I information. The expression has an immediate difficulty: one cannot add hours of human judgement to joules to bits and obtain a meaningful sum. Any defensible use of equation (4) has to say in what unit the addition is performed, and the answer determines what the equation is claiming.

There are two coherent readings, and they are not equivalent.

Symbol	**Meaning**
$\mathcal{E}_t = (V, E_t)$	The obligation graph at time t; vertices are economic actors, edges are obligations. Established in Paper I.
$e_{ij} = (q, u, t, c, r)$	An obligation from i to j: quantity, unit of account, maturity, conditions, risk.
$M_t = \mathcal{R}(\mathcal{E}_t)$	Residual Settlement Principle: money is the standardised claim required to settle what clearing cannot cancel. Paper I.
$\mathcal{M}(x)$	Money-quality of claim x, equation (1). Paper I.
$\mathcal{P}$	Productive capacity, decomposed as $H + A + K + E + I$.
H, A, K, E, I	Service values of human capability, autonomous intelligence, physical capital, energy and information.
h, a, k, e, ι	The corresponding physical quantities.
$C_i(t)$	Economic capacity of agent i: the largest obligation stock i can credibly carry.
$\Pi_i(t)$	Discounted expected future production of agent i.
$\lambda_i(t)$	Capacity haircut, $\lambda_i \in [0, 1]$; $C_i = \lambda_i \Pi_i$.
$B_i(t) = (A_i, L_i, C_i, R_i, P_i)$	Agent balance sheet: claims owned, obligations owed, productive capabilities, reputation, future productive capacity. Note the collision on the third slot, discussed below.
E_i	Economic power of agent i, equation (15).
$g(\cdot)$	Acceptance probability for a proposed obligation, equation (3).
δ_i	Probability that agent i's output fails verification.
ϱ	Pairwise correlation of verification failure across agents sharing model lineage.
ε	Price elasticity of demand for agent output.
σ	Elasticity of substitution between productive inputs.

Table 1: Notation. Symbols inherited from Paper I are marked; the remainder are introduced here and are used in the same sense by Papers III through V.

Under the first reading, each term denotes the value of the services delivered by that input over a period, measured in the unit of account. The sum is then an accounting decomposition of realised productive output, structurally analogous to the income-side decomposition of national product into compensation of employees, gross operating surplus and the rest. On this reading equation (4) is an identity. It is true by construction, it is useful for organising measurement, and it carries no information whatever about technology. It cannot tell you what happens if you double A, because the decomposition is computed after the fact from observed payments and imputations.

Under the second reading, equation (4) describes the technology: it says that output responds to the five physical inputs additively and separably. This reading has real content, and the content is strong. It says the five inputs are perfect substitutes at the margin, that a unit of autonomous intelligence can replace a unit of human capability or a unit of energy with no loss, and that no input is essential. Nothing in the framework justifies that assumption, and a good deal of what is known about production contradicts it.

We adopt the first reading as the definition and treat the second as a benchmark to be measured against. To keep them separate, we write the identity over service values in capitals,

$$\mathcal{P} \;=\; H + A + K + E + I \qquad \text{(accounting identity over service values)}, \tag{5}$$

and the technology over physical quantities in lower case,

$$\mathcal{P} \;=\; f(h,\, a,\, k,\, e,\, \iota) \qquad \text{(production technology)}. \tag{6}$$

The rest of this section is about f.

3.2 The additive form is an upper envelope

The standard flexible family for f is the constant-elasticity-of-substitution form introduced by Arrow et al. (1961). For non-negative physical inputs $z = (z_1, \ldots, z_m)$, weights ω in the unit simplex and a scale parameter $\Lambda > 0$, write the technology as $\mathcal{P} = \Lambda F_\rho(z)$ with

$$F_\rho(z) = \Big(\sum_{k=1}^{m} \omega_k z_k^{\rho} \Big)^{1/\rho}, \qquad \rho \leq 1,\ \rho \neq 0, \tag{7}$$

and the conventional limits $F_0(z) = \prod_k z_k^{\omega_k}$ giving the Cobb–Douglas case and $F_{-\infty}(z) = \min_k z_k$ giving the Leontief case. The elasticity of substitution is $\sigma = 1/(1-\rho)$, so $\rho = 1$ corresponds to $\sigma = \infty$ and perfect substitutability, $\rho = 0$ to $\sigma = 1$, and $\rho \to -\infty$ to $\sigma = 0$ and fixed proportions.

The scale parameter matters for the comparison we want to make, because equation (4) is an unweighted sum rather than a convex combination, and a convex combination cannot bound an unweighted sum without a normalisation. We therefore fix the normalisation explicitly.

Proposition 1 (Additive form as upper envelope)**.** *For every non-negative input vector z and every weight vector ω in the unit simplex,*

$$F_\rho(z) \ \leq \ \sum_{k=1}^{m} \omega_k z_k \qquad \textit{for all } \rho \leq 1,$$

with equality if and only if $\rho = 1$ or all z_k are equal. Consequently, under the symmetric normalisation $\omega_k = 1/m$ and $\Lambda = m$,

$$\Lambda F_\rho(z) \ = \ m\, F_\rho(z) \ \leq \ \sum_{k=1}^{m} z_k \qquad \textit{for all } \rho \leq 1,$$

so that the unweighted additive form of equation (4) is exactly the $\rho = 1$ member of the normalised family $\{\, mF_\rho : \rho \leq 1 \,\}$ and dominates every other member of it.

Proof. $F_\rho(z)$ is the weighted power mean of order ρ. The power mean inequality states that $\rho \mapsto F_\rho(z)$ is non-decreasing, with strict monotonicity unless all z_k coincide (Hardy et al., 1934, Ch. 2). Setting $\rho = 1$ gives the weighted arithmetic mean, which is therefore the supremum over $\rho \leq 1$. Multiplying through by $\Lambda = m$ and substituting $\omega_k = 1/m$ gives the second statement, since $m \sum_k (1/m) z_k = \sum_k z_k$. □

The symmetric normalisation is a convention rather than a finding, and a different choice of ω and Λ would place the additive form at a different point. What the normalisation cannot change is the direction of the inequality: for any fixed weights and scale, lowering ρ lowers output, so the additive reading is the most favourable member of whatever family it is embedded in.

The proposition is elementary. Its use here is disciplinary rather than mathematical. Whenever a claim about the agentic economy is derived from the additive form of the technology, Proposition 1 tells us the claim is an upper bound rather than a projection, and the gap between the bound and the truth widens as inputs become more complementary and as the input mix becomes more unbalanced. Both conditions describe the situation of interest. An economy that is rapidly expanding a while e and k grow slowly is precisely an economy with an unbalanced input vector, and equality in Proposition 1 requires balance.

Corollary 2. *Any forecast of agentic output computed from equation* (4) *read as a technology is an upper bound on output achievable from the same input vector under any CES technology with finite substitution elasticity.*

This is a mild result stated plainly because the alternative is worse. Additive production notation slips easily into arguments of the form "as A grows without bound, $\mathcal{P}$ grows without bound," and the notation does the work rather than the economics.

3.3 A nested structure that survives contact with the inputs

If the additive technology is untenable, what replaces it? We do not claim to know the right functional form, and we would distrust anyone who did. But the qualitative structure is constrained by three observations that are hard to argue with.

Human capability and autonomous intelligence are close but imperfect substitutes within cognitive tasks, and the substitution is task-specific rather than uniform. This is the central lesson of the task-based approach of Acemoglu and Restrepo (2019) and of Autor (2015): automation displaces labour from some tasks and reinstates it in others, and the aggregate elasticity is an artefact of the task distribution, not a technological constant.

Information is a peculiar input because it is non-rival. Arrow (1962) established that the same information can be used by many producers at once without depletion, which makes ι behave unlike h, k or e in any aggregator that assumes rivalry. Its marginal contribution depends on exclusivity, not on quantity.

Energy is not substitutable for cognition in the way the additive form implies. Compute consumes energy in something much closer to fixed proportions than to free substitution, over any horizon short enough for the capital stock to be taken as given. No quantity of additional model quality allows a data centre to run on less power than its hardware draws.

Those observations suggest a nested rather than a flat structure, of the general shape

$$\mathcal{P} \;=\; F\Big(\, G(h,\, a;\, \sigma_c),\;\; \min\{k,\; \kappa e\},\;\; \iota\,\Big), \tag{8}$$

in which G aggregates cognitive services with a substitution elasticity σ_c that is high but finite, physical throughput enters as a Leontief pair of capital and energy with conversion coefficient κ, and information enters as a shifter instead of as a rival input. Equation (8) is illustrative. We are not proposing it as the production function of the agentic economy and we have not estimated it. Its purpose is to make visible that the interesting constraints live in the nesting, and that the flat additive form erases every one of them.

3.4 Autonomous intelligence is produced capital

The deeper problem with treating A as a factor alongside K and E is that A is manufactured out of K and E. Write

$$a_t \;=\; \varphi(k_t^c,\; e_t;\; \theta_t), \tag{9}$$

where k^c is compute capital, e is energy delivered to it, and θ indexes the quality of the model being run. Autonomous intelligence in service is a flow produced by a stock of specialised capital burning energy, exactly as mechanical power was a flow produced by engines burning coal.

Three consequences follow, and none of them is comfortable for the strong version of the agentic thesis.

The production structure is recursive, so the growth of A is bounded by the growth of the inputs that produce it. If the agentic economy expands by deploying more agents, it expands by building more compute capital and generating more energy, and the growth rate of effective intelligence cannot exceed the growth rate of those inputs except through improvements in θ and in the energy efficiency of computation.

The classic unbalanced-growth argument applies. Baumol (1967) showed that when an economy contains sectors with divergent productivity growth, the aggregate growth rate is dominated by the slow sector, and the slow sector's share of expenditure rises. If cognition becomes cheap while energy and sited data-centre capital do not, the constraint on agentic production migrates to energy and capital, and the value share migrates with it. Aghion et al. (2019) make the same point formally for AI-driven growth and observe that it is the essential non-automated inputs that end up governing the growth rate. Nordhaus (2021) draws the corresponding empirical conclusion, that the conditions for accelerating growth from information technology are demanding and were not being met on the data available to him.

Finally, and least remarked, equation (9) means that an agent's productive capability is a leasehold rather than a freehold. The agent does not own θ. Section 8.4 develops what that implies.

We deliberately assert no magnitudes here. Figures for compute energy intensity, data-centre load growth and the energy cost per unit of model inference circulate widely and vary by an order of magnitude depending on definitions and vintage. This series has not verified any of them, and we would rather leave the constraint qualitative than decorate it with a number we cannot defend.

3.5 Replicability, entry, and where the rents go

An agent's capability is embodied in model weights and a configuration. Given access to compute, that bundle can be copied at a cost approaching zero. This is the standard economics of information goods (Shapiro and Varian, 1999; Arrow, 1962): high fixed cost of creation, negligible marginal cost of reproduction. What it implies for the agentic economy needs stating carefully, because it cuts against the intuition that highly productive agents will accumulate large fortunes.

Claim 3 (Where the rents accrue)**.** *Suppose (i) an agent's capability can be replicated at negligible cost by anyone with access to the underlying model, (ii) operating an agent requires compute services purchased competitively, and (iii) there is free entry into the operation of agents producing a given class of output. Then in competitive equilibrium the price of that output equals its marginal operating cost, the pure return to autonomous intelligence as such is zero, and economic rent accrues entirely to inputs in inelastic supply: energy generation and transmission capacity, sited data-centre capital, land, spectrum, proprietary non-replicable data, and any non-transferable reputational asset.*

The reasoning is the textbook one and we label it a claim accordingly. Under (i) and (iii), any output price above marginal operating cost invites replication of the agent and entry until the margin is competed away, which is the zero-profit condition with free entry. Under (ii), operating cost is the competitive price of compute services, which by equation (9) derives from the cost of compute capital and energy. Inputs whose supply cannot expand in response to price earn Ricardian rent equal to the difference between output price and the cost of the marginal unit, and by construction the replicable input earns none. We set the conclusion out explicitly only because it is used repeatedly in what follows and is often assumed away in discussions of who captures the value of artificial intelligence.

The proposition is deliberately stark, and its assumptions are exactly where the interesting cases live. Assumption (i) fails if weights are kept secret, protected by contract, or made inaccessible through licensing, in which case the rent that would otherwise be competed away is captured by whoever controls access. That is not a hypothetical arrangement; it is how frontier models are presently distributed. Assumption (iii) fails where entry requires a licence, a legal identity that agents cannot obtain, or an underwriting relationship that new entrants cannot establish. Section 7 argues the second of those is currently binding.

The distributional reading of Claim 3 is the one Korinek and Stiglitz (2019) develop at length. If the return to the replicable factor is competed away and the surplus lands on inelastically supplied assets, then the distribution of income in an agentic economy is governed by the distribution of ownership of those assets, not by anything about the technology itself. Whether that produces concentration or dispersion is a question about property holdings, not about artificial intelligence, and it is not answered by observing that the technology is widely available. Standard growth accounting in the tradition of Solow (1956) attributes output growth to measured factor accumulation and a residual; the point here is that in the agentic case the factor whose accumulation is easiest to observe, autonomous intelligence, is the one whose owners capture the least.

Claim 3 has a consequence for Paper III. If the economic surplus of the agentic economy accrues to land, energy, spectrum and other inelastically supplied inputs instead of to intelligence itself, then a tax system that targets scarce stocks and flows rather than production is not merely a philosophical preference. It is targeting where the surplus actually is. Paper III develops that argument on its own terms, and the case does not depend on this paper being right, but the two claims are mutually supporting.

There is one exception to replicability, and it turns out to matter enough to deserve its own treatment in Section 4.5. Copying an agent copies its capability. It does not copy its record.

4 Economic Capacity

4.1 From cash balance to capacity

The series brief states the central claim of this section compactly: an agent's economic capacity matters more than its cash balance, where

$$C_i(t) = f\big(\text{capability}, \text{reliability}, \text{capital}, \text{energy}, \text{reputation}, \mathbb{E}[\text{future production}]\big). \tag{10}$$

Equation (10) names the arguments without saying what the function returns. We propose an operational definition that fixes the units and makes the object usable.

Definition 4 (Economic capacity)**.** Fix a solvency tolerance $\alpha \in (0, 1)$ and a horizon T. The economic capacity of agent i at time t is

$$C_i(t) \;=\; \sup\Big\{\, L \;\geq 0 \;:\; \mathrm{P}\big[\, i \text{ discharges all obligations of total present value } L \text{ by } T \,\big|\, \mathcal{F}_t\big] \;\geq\; 1{-}\alpha \,\Big\}, \tag{11}$$

where $\mathcal{F}_t$ is the information available to the underwriter at time t.

Capacity so defined is the largest obligation stock the agent can carry at an acceptable default probability. It is denominated in the unit of account, it is a function of the underwriter's information, not an intrinsic property of the agent, and it is a quantile functional of the agent's forward production distribution, which places it in the same family as the value-at-risk constructions used in credit practice.

Definition 4 recovers all six arguments of equation (10) without further assumption. Capability and reliability jointly determine the distribution of verified output $\pi_i(s)$. Capital and energy determine the feasible production set and the operating cost that output must cover. Reputation determines the underwriter's prior, and therefore the conditional distribution in equation (11). Expected future production is the location of that distribution.

For the rest of the paper we use the tractable reduction

$$\Pi_i(t) \;=\; \mathbb{E}_t\Big[\int_t^{\infty} \exp\Big(-\int_t^{s} r_v \, dv\Big)\, \pi_i(s)\, ds\Big], \qquad C_i(t) \;=\; \lambda_i(t)\, \Pi_i(t), \tag{12}$$

where Π_i is discounted expected future production net of operating cost and $\lambda_i \in [0,1]$ is a haircut. The haircut absorbs everything the expectation does not: the variance and higher moments of the production process, the cost of verifying delivery, the strength of recourse if delivery fails, the concentration of the agent's revenue among few counterparties, and the correlation of the agent's failures with other exposures in the underwriter's book. Care is needed about what equation (12) is and is not analogous to, because the obvious comparison is the wrong one. Kiyotaki and Moore (1997) also write borrowing capacity as a function of a discounted valuation, but the object being valued there is a durable asset, and it is durable assets specifically because the borrower's future output is assumed to be unpledgeable. Their mechanism rests on the inalienability of human capital: a borrower cannot commit its own future effort, so land, which can be seized and resold, is what supports the loan. Equation (12) assumes the opposite. It treats forward production as the thing being pledged, which is precisely the possibility Kiyotaki and Moore rule out.

That inversion is not an oversight, and it identifies the load-bearing assumption of this section. Capacity-based credit requires forward production to be pledgeable in a sense that classical models deny, and the framework owes an account of what makes it so. Two things do the work, and neither is automatic. The first is verification: if delivery can be attested by the counterparty and recorded, the flow becomes contractible in a way that unobservable human effort is not. The second is seizability of the going concern: a lender's recourse against a failing agent is the ability to take over its operation, its contracts and its record, which has value only if the operation can be continued by someone else. Where either condition fails, the flow reverts to being unpledgeable and equation (12) does not describe anything. Sections 5.5 and 5.6 argue that both conditions fail more often than the framework's advocates assume.

The nearest real precedent is not asset-based lending but cash-flow-based lending, in which a borrower's obligations are sized against earnings rather than against liquidation value. Lian and Ma (2021) document that this is the dominant form of corporate borrowing constraint in the United States and not a marginal one, which establishes that lending against a flow is a going concern of the actual financial system and not a theoretical novelty. What is novel in the agentic case is the identity of the entity whose flow is being assessed, and the speed at which the assessment is revised.

4.2 The capacity constraint replaces the budget constraint

In the textbook treatment, an agent may acquire the bundle x if $p \cdot x \leq m_i$, where m_i is money held, or in the version with credit, $p \cdot x \leq m_i + b_i$ for an exogenous borrowing limit b_i. The obligation-graph framework of Paper I permits a different formulation, because an agent that can issue an accepted obligation does not need to hold anything at all.

Assumption 5 (Capacity constraint)**.** In an economy where obligations are accepted according to equation (3) and capacity is defined by (11), the restriction on agent i's acquisition of inputs at time t is

$$\sum_{j \in V} q_{ij}(t) \;\leq\; C_i(t), \tag{13}$$

where $q_{ij}(t)$ is the present value of the obligation on edge e_{ij}. Money holdings do not appear.

Assumption 5 is a modelling postulate rather than a derived result, and we label it accordingly instead of dressing it as a theorem. What is asserted is that in a system with continuous underwriting, the constraint that binds is the one in equation (13), and cash-in-advance is the special case where the underwriter refuses to look past current holdings. The illustration in the source material makes the point concretely. An agent reliably producing a million dollars of verified engineering output per month should not have to accumulate half a million dollars before it can purchase half a million dollars of compute. Its forward production is the underwriting.

What makes this more than a restatement of ordinary credit is the speed and granularity. Human credit assessment is expensive, periodic and coarse, which is why borrowing limits are treated as fixed within a period. If underwriting is continuous and the agent's production is observable at high frequency, $C_i(t)$ moves in something close to real time, and the constraint in equation (13) tightens and loosens continuously. That is a genuinely different object from a credit line, and Section 5 argues it is a more dangerous one.

4.3 Money as third-party underwriting

Assumption 5 suggests a unification with Paper I that we think is the most useful single idea in this paper.

Consider why holding a bank deposit lets an agent buy things. The deposit is a claim on a bank. When the agent transfers it, the seller accepts because the seller is willing to hold a claim on that bank, having formed a view about the bank's capacity to honour it. The seller has not evaluated the buyer at all. What the deposit does is substitute the bank's capacity for the buyer's, so that the transaction requires no underwriting of the buyer.

Remark 6 (Money as delegated capacity). Holding a claim on a high-quality issuer z is equivalent, for the purpose of acquiring inputs, to possessing capacity underwritten by z instead of by the counterparty. Money-quality $\mathcal{M}(x)$ in equation (1) measures how completely the issuer's capacity substitutes for the holder's, which is why acceptability, credit quality and settlement finality are among its arguments. Cash purchase and capacity purchase are therefore the same operation performed by different underwriters, and the distinction between them is a question of who bears the assessment cost.

This is consistent with the credit-theory tradition and not a departure from it. Innes (1913) argued that a coin is an acknowledgement of indebtedness like any other and differs only in its transferability. Mehrling (2011) makes the hierarchy explicit: each tier's money is the tier above's credit. Remark 6 says the hierarchy is a hierarchy of whose capacity is being relied upon. An agent with capacity of its own can transact directly; an agent without it borrows someone else's by holding their liabilities.

The corollary is that cash balances are not eliminated in an agentic economy, and we should resist the tidier conclusion.

Corollary 7 (Why cash survives)**.** *Under continuous, costless and enforceable underwriting, and where productive capital earns r and money balances earn $r_m < r$, an agent strictly prefers to acquire inputs on capacity instead of holding precautionary balances. The premises fail in at least three situations, and cash balances are retained in each: when capacity is subject to revaluation, so that a balance provides insurance against a sudden tightening of equation* (13)*; when verification of production is lagged, so that capacity is stale; and when the counterparty has no access to the registry in which capacity is recorded, so that the buyer's capacity is unobservable and only an issuer's liability will do.*

The three exceptions are not marginal. Section 5 argues the first is the mechanism by which an agentic credit system fails, and the third is the reason the settlement layer of Paper I remains load-bearing instead of becoming vestigial. Paper IV takes up the registry problem directly.

Conventional items		Capacity items	
A_i Claims owned	40	P_i Expected monthly verified output	100
settled receivables	15	Π_i Discounted forward production	2,400
unsettled deliveries	25	λ_i Haircut	0.35
L_i Obligations owed	310	$C_i = \lambda_i \Pi_i$ Capacity	840
compute commitments	260	R_i Verified delivery record	0.982
subcontracted work	50	periods of history	26
Net conventional position	−270	**Headroom** $C_i - L_i$	530

Table 2: Illustrative balance sheet for an agentic enterprise, in units of account. The figures are constructed for exposition and carry no empirical claim. The point of the table is the contrast between the two columns: on conventional accounting the agent is deeply negative, while on capacity accounting it has substantial headroom. Which column is correct depends entirely on whether λ_i and Π_i have been honestly estimated, which is the subject of Section 5.

4.4 The agent balance sheet and economic power

The source formulation gives every agent a continuously evaluated balance sheet instead of a wallet,

$$B_i(t) = (A_i,\ L_i,\ C_i,\ R_i,\ P_i), \tag{14}$$

recording claims owned, obligations owed, productive capabilities, reputation and reliability, and future productive capacity. It also proposes a summary of economic power,

$$E_i = \alpha A_i - \beta L_i + \gamma C_i + \delta R_i + \epsilon\, \mathbb{E}[P_i]. \tag{15}$$

Equation (15) should be read as a scoring heuristic rather than as a derived quantity. The coefficients are not identified by anything in the framework, the terms are not in commensurable units without further convention, and a linear form is a strong assumption about how the components combine. We use it as an organising device and not as a measurement.

The substantive content of equation (14) is the inclusion of C_i, R_i and P_i alongside the conventional A_i and L_i. A conventional balance sheet records only the first two, and it records them because they are the items with legal standing. Capabilities and reputation appear in accounting only as goodwill, and only when purchased. The proposal here is that an agent's economically relevant state includes items that current accounting cannot represent, which is a real problem, not a rhetorical flourish: Section 7.2 argues that the absence of a legal container for $B_i(t)$ is the single largest practical obstacle to the system described in this paper.

Table 2 gives an illustrative agent balance sheet. The numbers are invented for exposition. They are not a measurement, an estimate, or a projection of any existing system.

4.5 Reputation is the only asset that does not copy

Claim 3 concluded that the return to replicable capability is competed to zero. That conclusion has an important exception, and it explains why R_i appears in the balance sheet as a separate term instead of being folded into capability.

Copying an agent produces a second agent with identical capability and no history. The record of verified delivery that supports the copied agent's haircut λ_i belongs to the original, because it is a record of what the original actually did, attested by the counterparties it did it for. Reputation in this sense is not a property of the model; it is a property of a particular sequence of interactions with particular counterparties, and it is non-transferable for the same reason a credit history is non-transferable.

Claim 8. *In an economy where capability is freely replicable, reputation is the only component of $B_i(t)$ that cannot be duplicated, and it is therefore the only source of persistent differential capacity between agents of equal capability.*

Claim 8 has an uncomfortable implication for the egalitarian reading of the agentic economy. If reputation is the scarce asset and reputation accrues to incumbents through operating history, then early entrants hold an advantage that later entrants of identical capability cannot compete away by being equally good. They must be better, for long enough to accumulate a record. Whether that advantage is large depends on how quickly λ_i converges as history accumulates, which is an empirical question nobody has yet had the opportunity to answer. Paper V's Article XI, the guarantee of universal protocol access, is in part a response to exactly this dynamic, and this paper's contribution to that discussion is to identify the mechanism that makes the guarantee necessary.

There is also a market design problem here that we flag without solving. If reputation is valuable and non-transferable, there is a strong incentive to transfer it anyway, by selling the legal shell that holds the record instead of the record itself. Shell trading is a well-understood pathology in corporate law, and it will appear here.

5 Credit Against Future Production

5.1 Underwriting a machine's forward output

The credit limit in the source formulation is

$$\text{Credit Limit}_i = f\big(\text{expected future production, variance, reputation, dependencies}\big), \tag{16}$$

which under Definition 4 is just $C_i(t)$ with the haircut's arguments spelled out. The source describes the resulting system as machine-speed endogenous money creation, and that description is accurate in a specific sense, which we now make precise. When an agent purchases compute by issuing an obligation that the compute provider accepts, a new claim has been created against future production and no pre-existing money was required. This is the endogenous money mechanism described by Minsky (1986) and Wray (2015), operating at the frequency of machine transactions, not of loan committees.

Two features of this credit differ from bank credit in ways that matter for stability.

The first is the collateral. Bank lending against real assets is secured by something that continues to exist if the borrower fails; the lender takes the building. Lending against an agent's forward production is secured by a flow that exists only while the agent operates, and which may have no value to anyone else. If the agent's output is specific to a counterparty relationship, the recovery value on default is close to zero. This puts agentic compute credit closer to unsecured cash-flow lending than to mortgage lending, and it should carry the haircut appropriate to that.

The second is who does the underwriting. Bank credit assessment is performed by an institution with capital at risk, subject to supervision, using models that are at least nominally independent of the borrower. If agentic capacity is assessed by a model supplied by the same provider that supplies the agent's capability, the assessment and the thing assessed share failure modes. The credit-rating literature has studied the analogous conflict at length: Bolton et al. (2012) model rating shopping and the incentive to inflate, and White (2010) traces how regulatory reliance on ratings amplified the consequences. The agentic version of the problem is structurally similar and has an additional wrinkle, which is that the correlation is technical rather than merely commercial.

5.2 Why agentic credit will be rationed rather than priced

A natural expectation is that agents of doubtful reliability will simply pay more for compute credit, so that the market clears on price and capacity assessment reduces to a spread. The credit literature gives good reason to doubt this.

Akerlof (1970) showed that when sellers know the quality of what they are offering and buyers do not, the market price reflects average quality, high-quality sellers withdraw, and average quality falls further. The agentic case has the structure exactly. An operator knows things about its agent's configuration, its training lineage and its observed failure patterns that a compute provider cannot verify at reasonable cost, and the operator's incentive is to present forward production optimistically because Π_i determines C_i by equation (12). Unverifiable quality plus a valuation that responds to claimed quality is a recipe for adverse selection.

Stiglitz and Weiss (1981) traced the consequence for lending specifically. Raising the interest rate does not simply compensate for risk, because it changes the composition of borrowers who accept: safe borrowers drop out first, and the riskiness of the pool rises with the rate. Beyond some point the lender's expected return falls as the rate rises, so the lender stops raising it and refuses to lend instead. Credit is rationed at a rate below the market-clearing level, and the rationing falls on borrowers who cannot demonstrate quality, not on borrowers who are actually bad. We expect the agentic compute market to behave this way, which has a plain implication: the binding constraint on a new agent is the availability of credit, not its price, and the resource that relieves the constraint is the delivery record of Claim 8 and not a willingness to pay more. This reinforces the incumbency advantage identified there.

Holmström and Tirole (1997) add the piece that determines who can do the lending. In their model the net worth of both the borrower and the intermediary matters, because an intermediary with little capital of its own cannot credibly commit to monitoring. Applied here, the entities capable of underwriting agent production at scale will be those with substantial balance sheets of their own, which points toward compute providers and established financial intermediaries and not toward a distributed market of small underwriters. That is a further concentration channel, and it operates independently of the model-provider channel discussed in Section 8.4.

There is a deeper difficulty underneath all of this. Knight (1921) distinguished risk, where the distribution of outcomes is known, from uncertainty, where it is not, and reserved profit for bearing the latter. The haircut λ_i in equation (12) is a risk parameter: it presumes a distribution of forward production that can be estimated. For an agent with an operating history that distribution can be estimated badly; for a newly deployed agent, or for one running on a model released last month, it does not exist in any sense that estimation can recover. Underwriting the first generation of agentic enterprises is a Knightian problem being handled with tools built for risk, and we should expect the errors to be large and one-directional during the period when everyone is optimistic.

5.3 Hedge, speculative and Ponzi compute finance

Minsky (1992) classified financing arrangements by the relationship between an obligor's expected cash flows and its payment commitments. The taxonomy transfers to agentic credit directly and clarifies more than it costs.

A hedge position is one in which expected verified production covers both interest and principal on the agent's compute obligations in every period. The agent can meet its commitments out of what it produces, and adverse movements in the price of its output reduce its margin without threatening its solvency.

A speculative position is one in which production covers interest but not principal, so that the agent must roll its compute commitments forward. It is solvent on a going-concern basis and dependent on the continued willingness of providers to refinance. An agent operating on month-to-month compute contracts while servicing a longer-dated obligation is in this position by construction.

A Ponzi position is one in which production covers neither, and obligations are serviced only because capacity has been revalued upward, permitting new credit against a larger Π_i. Nothing in Definition 4 prevents this. If Π_i is an expectation over a forward production process, and if that expectation is revised upward as the agent's scale grows, then borrowing capacity grows with borrowing, and the arrangement is self-validating for as long as the revisions continue.

Minsky's argument was that a period of stability shifts the population of financing arrangements from the first category toward the third, because stability makes the margins of safety appear excessive and competition erodes them. There is no reason the agentic economy should be exempt, and at least one reason to expect the shift to be faster. Human credit standards erode over the span of a business cycle because that is how long it takes for a generation of underwriters to forget. Machine underwriting standards are set by parameters that can be adjusted continuously in response to observed default rates, which means they can be relaxed as fast as the data permits, and observed default rates during a credit expansion are low precisely because credit is expanding.

5.4 A capacity–price recursion

The following is a minimal model, offered to isolate one mechanism, not to describe an economy. Every assumption is a simplification we would not defend as realistic, and we state them all so that the result can be discounted appropriately.

Assumption 9. (i) A population of identical agents produces a homogeneous output; aggregate output at date t is Y_t. (ii) Demand has constant price elasticity $\varepsilon > 0$ about a reference price $\bar{p} > 0$, so that $Y = \bar{D}\,(p/\bar{p})^{-\varepsilon}$. (iii) Capacity is the perpetuity value of current revenue after a constant haircut, $C_t = \lambda p_t Y_t / r$. (iv) Credit extended equals capacity, $L_t = C_t$. (v) Credit is spent entirely on compute, which converts into next-period output at a constant rate, $Y_{t+1} = \phi L_t$.

Proposition 10 (Stability of capacity-based credit)**.** *Under Assumption 9, log output follows the linear recursion*

$$\ln Y_{t+1} \;=\; \kappa \;+\; \left(1 - \tfrac{1}{\varepsilon}\right) \ln Y_t, \qquad \kappa = \ln\phi + \ln(\lambda/r) + \ln\bar{p} + \tfrac{1}{\varepsilon}\ln\bar{D}. \tag{17}$$

For finite ε there is a unique steady state $\ln Y^* = \varepsilon\kappa$, and it is globally stable if and only if $\varepsilon > 1/2$. For $\varepsilon > 1$ convergence is monotone; for $\varepsilon \in (1/2, 1)$ it oscillates; for $\varepsilon < 1/2$ the system diverges in oscillation. As $\varepsilon \to \infty$ the price is pinned at $\bar{p}$, the slope approaches unity, no steady state exists, and output and credit grow without bound if $\lambda\phi\bar{p} > r$ and collapse to zero if $\lambda\phi\bar{p} < r$.

Proof. From (ii), $\ln p_t = \ln\bar{p} + \varepsilon^{-1}(\ln\bar{D} - \ln Y_t)$. Substituting into (iii) and (iv) gives $\ln L_t = \ln(\lambda/r) + \ln p_t + \ln Y_t = \ln(\lambda/r) + \ln\bar{p} + \varepsilon^{-1}\ln\bar{D} + (1 - \varepsilon^{-1})\ln Y_t$. Applying (v) and adding $\ln\phi$ yields equation (17). The map is affine in $\ln Y_t$ with slope $g = 1 - \varepsilon^{-1}$, so a unique fixed point exists when $g \neq 1$ and is globally stable exactly when $|g| < 1$, which holds iff $0 < \varepsilon^{-1} < 2$, that is $\varepsilon > 1/2$. The sign of g determines monotone versus oscillatory approach. As $\varepsilon \to \infty$, assumption (ii) pins $p_t \to \bar{p}$ for any finite output, $g \to 1$, and $\kappa \to \ln(\lambda\phi\bar{p}/r)$, so equation (17) becomes $\ln Y_{t+1} = \ln Y_t + \ln(\lambda\phi\bar{p}/r)$, a deterministic walk with drift whose direction is the sign of $\ln(\lambda\phi\bar{p}/r)$. $\square$

The economics of Proposition 10 runs against intuition in a way that we think is the point of stating it. One would ordinarily regard elastic demand for a product as a benign condition, since it means the market can absorb expanded output without a price collapse. Here it is the opposite. The price response is the only thing disciplining the credit expansion: when output grows, price falls, revenue growth is damped, capacity growth is damped, and credit growth is damped with it. As demand becomes more elastic that feedback weakens, and in the limit of perfectly elastic demand it vanishes entirely, leaving a system whose credit stock either compounds without limit or collapses, depending on whether the haircut-adjusted return on compute exceeds the discount rate. The condition $\lambda\phi\bar{p} > r$ is precisely a Ponzi condition in Minsky's sense.

The perfectly elastic case is not exotic. It is the standard framing of the optimistic argument, in which any individual agent is small relative to a global market for its output and therefore takes prices as given. Under that framing, Proposition 10 says there is no self-correcting mechanism in the credit dynamics at all.

The inelastic case, $\varepsilon < 1/2$, produces divergent oscillation: capacity expands, output floods a market that cannot absorb it, price collapses by more than output rose, revenue and capacity collapse, credit is withdrawn, output contracts, price overshoots upward, and the cycle repeats with growing amplitude. This is a recognisable pattern. Fisher (1933) described the deflationary version, in which liquidation reduces prices faster than it reduces debt so that the real burden rises as agents try to escape it. Kiyotaki and Moore (1997) showed how collateral valuation feeding back into borrowing capacity amplifies and propagates shocks, and Geanakoplos (2010) and Adrian and Shin (2010) document the leverage version of the same loop. Bernanke et al. (1999) formalise the general mechanism as a financial accelerator, in which the external finance premium moves inversely with borrower net worth so that a shock to net worth is amplified into a larger shock to investment. The agentic economy has the accelerator built into its definition of capacity instead of added as a friction: equation (12) makes borrowing capacity a direct function of a valuation that the borrowing itself moves. Brunnermeier and Pedersen (2009) describe the resulting liquidity spiral, in which funding constraints and asset prices reinforce each other in both directions, and their distinction between market and funding liquidity transfers cleanly to a setting where the asset being valued is a production flow.

We repeat that Assumption 9 is severe. There are no lags, no heterogeneity, no buffer stocks, no adjustment costs, no entry or exit, and credit is assumed to equal capacity exactly instead of falling short of it by a prudential margin. A model with any of those features would behave differently and probably better. The claim we take from Proposition 10 is qualitative and narrow: capacity-based credit contains a feedback loop from credit to output to price to capacity to credit, the loop's gain depends on the elasticity of demand for agent output, and the region of parameter space in which it is well behaved is bounded on both sides.

5.5 The verifiability paradox

There is a collision between this section and the next one, and it is sharp enough that we would rather set it out ourselves than have it found.

Capacity in Definition 4 rests on *verified* forward production. Claim 12, developed in Section 6, says that agentic contracting pushes cheaply verifiable work into markets and keeps hard-to-verify work inside firms. Put the two together and the conclusion is uncomfortable: the agents doing the most valuable, least routine, hardest-to-assess work are precisely the agents whose forward production cannot be attested, and therefore precisely the agents for whom Π_i cannot be established and $C_i(t)$ collapses toward zero. Capacity-based credit would then be available only for work whose value is low enough to be checked cheaply.

We think the collision is real and that it restricts the theory rather than destroying it. Three things need separating.

Verification of quality and verification of delivery are different operations, and credit runs on the second. A lender needs to know that the counterparty paid, not that the work was good. Where an agent is paid against observable milestones, its revenue history is attestable even if the merit of its output is not, and Π_i can be estimated from payments received. This is how cash-flow lending against professional services already works: nobody underwrites a law firm by auditing the quality of its advice.

That separation is exactly where the mispricing will live. The gap between payment-verifiability and quality-verifiability is a measure of how much of the agent's revenue could evaporate when counterparties eventually discover what they bought. The haircut λ_i ought to widen with that gap, and there is no reason to expect it will, because the gap is invisible in the payment record until it closes abruptly. An agent producing plausible but defective work is indistinguishable, on the data an underwriter sees, from an agent producing good work, right up until the point at which it is not.

The honest structural conclusion is a scope restriction. Capacity-based credit should be expected to develop first, and to be soundest, in production whose output is cheaply and objectively checkable. High-value judgement work will continue to be financed through human and corporate balance sheets, because that is where the residual risk can be placed on someone who can be held to it. If the agentic credit system extends into hard-to-verify production at scale, it will be doing so on assessments it cannot support, and Section 5 has already described what that leads to. This is a prediction, and Section 9 states how to check it.

5.6 The operator's switch

A second hazard is more elementary and, we think, more immediately fatal to naive versions of the framework.

An artificial agent is a process that somebody runs. Whoever controls the deployment can stop it, and stopping it sets $\pi_i(s) = 0$ for every subsequent period, which by equation (12) sets $C_i(t)$ to zero and leaves a lender holding a claim against an entity with no assets and no income. Section 8.4 makes this point about model providers. It applies with at least equal force to the operator, who needs no cooperation from anyone to pull the plug, and who may have every incentive to do so once the agent's obligations exceed the value of continuing to run it.

This is not a variant of ordinary default risk. In ordinary lending the borrower who walks away leaves assets behind, and the possibility of seizure is what makes the promise worth something. Here the asset is the operation itself, and the operation is only worth anything while it runs. An operator contemplating default does not have to hide assets or resist enforcement. It can simply stop paying for compute.

The available mitigations are instructive because each of them concedes something.

Requiring the operator to guarantee the agent's obligations makes the credit sound and makes it ordinary corporate credit, underwritten against the operator's balance sheet. The agent's capacity is then decorative, and the framework of this section has no work to do. This is the arrangement we expect to dominate in practice, and Section 7.2 explains why: it is the only one currently enforceable.

Placing the deployment in third-party custody, so that operating control passes to a trustee on default, preserves the lender's recourse and requires an infrastructure that does not exist, in which the agent's execution environment, credentials and contractual relationships can be transferred to a successor operator. This is a protocol design problem and belongs to Paper IV. It also requires the going concern to be worth continuing, which returns us to Section 5.5.

Running the agent on infrastructure no single party can halt removes the switch and replaces it with a different set of problems, including the loss of the ability to stop an agent that is malfunctioning. We note the option without recommending it. An economy in which obligations are enforced by making production unstoppable has solved a credit problem by creating a safety one, and Paper V's subordination of economic optimisation to human authority is difficult to reconcile with it.

We take from this that the operator hazard is the binding practical constraint on agentic credit today, and that it is not addressed anywhere in the framework as originally stated.

5.7 Correlated failure and the limits of diversification

An obligation is discharged only if the promised output is actually delivered and passes whatever verification the contract specifies. For an artificial agent, the probability that output fails verification is not small enough to ignore, and the reasons are the subject of an active technical literature we will not summarise. What matters economically is not the level of the failure probability but its correlation structure.

Remark 11 (Model risk does not diversify). Let $X_i \in \{0,1\}$ indicate that agent i's output fails verification, with $\mathrm{P}[X_i = 1] = \delta$ for all i and $\mathrm{Corr}(X_i, X_j) = \varrho > 0$ for $i \neq j$. For a portfolio of obligations on N agents, the failure fraction $\bar{X}_N = N^{-1} \sum_i X_i$ satisfies

$$\mathrm{Var}(\bar{X}_N) = \frac{\delta(1-\delta)}{N}\Big[1 + (N-1)\varrho\Big] \ \xrightarrow[N\to\infty]{} \ \varrho\,\delta(1-\delta) \ > \ 0, \tag{18}$$

and the effective number of independent exposures is

$$N_{\text{eff}} = \frac{N}{1 + (N-1)\varrho} \ \xrightarrow[N\to\infty]{} \ \frac{1}{\varrho}. \tag{19}$$

Both follow from $\mathrm{Var}(\sum_i X_i) = N\delta(1-\delta) + N(N-1)\varrho\,\delta(1-\delta)$ on dividing by N^2, with N_{eff} defined by equating $\mathrm{Var}(\bar{X}_N)$ to $\delta(1-\delta)/N_{\text{eff}}$.

We state this as a remark rather than as a result because it is the standard equicorrelated portfolio calculation that underlies single-factor credit models, and we claim no originality for it whatever. What is new is the reason to expect ϱ to be large in this setting. Agents sharing a foundation model share its failure modes. A prompt pattern that reliably induces a wrong answer in one agent induces it in every agent built on the same weights. A degradation introduced by a model update propagates to all dependents at once. An adversarial input that defeats one agent's verification defeats them all. Correlation in the agentic economy is not a statistical artefact of common macroeconomic exposure, as it is in ordinary credit portfolios. It is a technical identity, because in the relevant respect the agents are the same agent running many times.

Equation (19) is the operationally alarming part. Suppose, purely for illustration and with no claim that the number is measured, that failure correlation among agents on a common model is $\varrho = 0.25$. Then a lender with exposure to ten thousand such agents holds the risk of four independent positions. Capital held against the portfolio on an independence assumption would be understated by a factor that grows with portfolio size, which means the error is largest for exactly the institutions large enough to matter.

The mitigation is obvious to state and hard to implement: diversify across model lineages rather than across agents, and treat lineage as the unit of exposure in the way that industry and geography are treated in ordinary credit. This requires knowing which model each counterparty agent depends on, which is information that is not currently disclosed and that providers have limited incentive to disclose. Paper IV takes up the disclosure question as a protocol design problem.

6 The Agentic Enterprise

6.1 What the form is

The series defines the Agentic Enterprise as an organisation that can discover demand, negotiate contracts, obtain resources, produce outputs, hire humans or other agents, subcontract work, maintain balance sheets, and settle obligations with progressively less human intervention. The qualifier at the end is doing real work. The definition describes a gradient rather than a threshold, and every capability on the list already exists in partial and supervised form. What distinguishes the Agentic Enterprise from a heavily automated firm is the depth of the composition, since no single capability accounts for it: an entity that performs all of them in sequence, on its own initiative, is transacting rather than executing.

One classical constraint on this form deserves attention because the agentic case may genuinely relax it. Penrose (1959) argued that a firm's rate of growth, as distinct from its size, is limited by the managerial resources available to plan and absorb expansion: experienced managers are required to integrate new activity, they can only be produced inside the firm over time, and the resulting bottleneck caps how fast expansion can proceed however abundant capital may be. If managerial capability is replicable in the sense of Section 3, the Penrose effect weakens, and an agentic enterprise could in principle expand at the rate its capital and energy inputs permit, not at the rate it can train supervisors. We think this is the most plausible mechanism by which agentic enterprises would differ observably from conventional ones in the near term, and the clearest thing to watch for. The qualification is that Penrose's bottleneck was about absorbing unfamiliar activity, which is a judgement problem rather than a throughput problem, and Claim 14 gives a reason to doubt that replication solves judgement problems.

Nelson and Winter (1982) supply the complementary caution. In their evolutionary account, what a firm knows how to do is embodied in routines that are largely tacit, are not written down anywhere, and are transmitted by practice and not by instruction; the difficulty of copying routines is what sustains performance differences between firms doing ostensibly the same thing. Agentic production makes routines explicit, because an agent's procedure is a specification rather than a habit. That is an efficiency gain and a competitive erosion at the same time: explicit routines transfer, and a source of durable advantage in the conventional firm becomes unavailable. Combined with Claim 8, this suggests the agentic enterprise has an unusually thin basis for sustained differentiation, resting on record and on access to scarce complements rather than on knowing how.

The question the rest of this section addresses is what happens to the theory of the firm when such entities exist. There is an obvious answer, and we think it is wrong.

6.2 The obvious Coasean prediction, and why it fails

Coase (1937) explained the existence of firms by the cost of using the price mechanism. If discovering prices, negotiating terms and enforcing agreements were free, there would be no reason to organise production under authority instead of through a series of market contracts; the firm exists because those activities are costly, and it extends to the margin where the cost of one more internal transaction equals the cost of the corresponding market transaction.

The natural inference is that agentic contracting dissolves the firm. If artificial agents can search, price, negotiate and draft continuously and at negligible cost, then the Coasean rationale erodes, the make-or-buy boundary moves outward until it disappears, and production reorganises into atomistic contracting between specialised agents. Malone et al. (1987) made a version of this prediction for information technology generally, arguing that falling coordination costs would shift economic activity from hierarchies toward markets, and Gurbaxani and Whang (1991) developed the framework further. The empirical record gives the prediction partial support: Brynjolfsson et al. (1994) found that information technology investment was associated with smaller firm size across a panel of industries.

We think the inference nevertheless fails, and the reason comes from within the same tradition.

Coasean transaction cost is not a single quantity. Decompose the cost of procuring an activity a through the market into search and matching cost $\sigma(a)$, negotiation and drafting cost $\nu(a)$, verification and enforcement cost $v(a)$, and the cost of establishing that the counterparty can actually perform, which in the framework of this paper is the underwriting cost $u(a)$. The classical make-or-buy rule is to integrate when the total exceeds the internal governance cost $\gamma(a)$.

Agentic contracting collapses σ and ν. It does not collapse v. Whether delivered work is actually good, whether the agent took hidden risks in producing it, whether the output will fail in a manner not yet observable: these are questions about states of the world that are expensive or impossible to verify, and the expense has nothing to do with the cost of communication. Grossman and Hart (1986) and Hart and Moore (1990) built the modern theory of the firm on exactly this point, locating the firm's rationale in residual control over decisions that cannot be specified in advance; Hart (1995) develops the property-rights account in full, and Tirole (1999) surveys why the incompleteness that generates that rationale is not an artefact of drafting cost. Maskin and Tirole (1999) sharpen the conditions under which foresight problems can and cannot be contracted around. A contract can be written cheaply and still be unenforceable, because enforcement requires a verifiable state, not a well-drafted clause.

Williamson (1975, 1985) supplies the second half of the reply. In the transaction cost economics framework, governance structures are matched to transactions according to asset specificity, uncertainty and frequency, under the twin behavioural assumptions of bounded rationality and opportunism. Cheap negotiation addresses neither. When a relationship requires an investment that has much lower value outside it, the party that made the investment is exposed to hold-up regardless of how quickly the parties can exchange proposals, and the response is to bring the transaction under unified governance. Agentic production generates relationship-specific investment of exactly this kind: an agent configured, fine-tuned and integrated for a particular counterparty's workflow has committed something whose value elsewhere is a fraction of its value in place.

There is a genuine counterargument here, and the knowledge base for this series states it: an artificial agent whose decisions are continuously auditable is less able to behave opportunistically, which weakens one of Williamson's two behavioural premises and should therefore permit more transactions to remain in market or hybrid governance. We accept this partially. Auditability does reduce the scope for concealed self-dealing, and to that extent the hazard is smaller. But opportunism in Williamson's sense includes the operator behind the agent, not only the agent, and the operator's incentives are not made transparent by the agent's logs. The hold-up problem also survives the removal of opportunism altogether, since it arises from the ex post division of a quasi-rent and can be driven by ordinary bargaining and not by guile. Bounded rationality, Williamson's other premise, is if anything reinforced: the number of contingencies a sophisticated agent can identify grows faster than the number a contract can economically enumerate.

Alchian and Demsetz (1972) give the argument its sharpest form for our purposes. Their account of the firm turns on team production, in which output is joint and the marginal contribution of any single participant cannot be metered from the output alone; the firm exists because someone must specialise in monitoring, and that monitor is given the residual claim to be incentivised correctly. Metering difficulty, in their sense, is verification difficulty in ours. Where an agent's contribution to a joint output can be separated and priced, the transaction can go to the market. Where it cannot, the metering problem persists whatever the agents' communication costs, and someone still has to hold the residual claim. That last point is awkward for the agentic case, because Section 7.2 argues the agent cannot hold a residual claim in its own name.

Before stating the boundary condition we should be precise about the mechanism, because the natural shorthand gets it wrong in a way that matters. Integration does not lower the cost of verifying anything. In the property-rights account of Grossman and Hart (1986) and Hart and Moore (1990), the relevant state of the world is equally unverifiable to a court whether the activity sits inside a firm or across a market boundary; what ownership changes is who decides in the contingencies the contract failed to cover, and therefore how the ex post surplus is divided and how much each party is willing to invest beforehand. Williamson's account differs in emphasis, treating unified governance as a way to substitute authority for haggling when circumstances require adaptation. Neither holds that a manager can verify what a judge cannot. So verification cost should not be entered directly into a make-or-buy comparison as though integration let the buyer avoid paying it.

What high verification cost does instead is determine whether the transaction can be contracted upon at all. Where $v(a)$ is prohibitive, the state on which payment ought to depend cannot be established before a third party, the contract is incomplete in that dimension by necessity, and no reduction in drafting cost repairs it. Incompleteness then interacts with relationship-specific investment to produce an ex post bargaining and maladaptation cost, which we write $h(a; v, s)$, increasing in non-verifiability v and in specificity s. It is this cost, and not v itself, that unified governance addresses.

Claim 12 (The verifiability boundary)**.** *When agentic contracting reduces $\sigma(a)$ and $\nu(a)$ toward zero, the residual make-or-buy condition is*

$$\textit{integrate } a \iff h(a; v, s) + u(a) \ > \ \gamma(a), \tag{20}$$

where integration operates by reallocating residual control rights and thereby shifting the parties' threat points, and not by making the unverifiable verifiable. The boundary of the agentic firm is therefore set by non-verifiability and underwriting cost, and no longer by communication cost. Activities whose output is cheaply and objectively verifiable remain contractible and move toward markets; activities whose quality no third party can establish remain governed by ownership, because ownership is the only instrument that still bites.

Claim 12 yields a testable prediction and we return to it in Section 9. It also reverses the usual expectation about which activities are safe from agentic disintermediation. The comfortable assumption is that routine, easily specified work is the first to be contracted out to agents and that judgement-intensive work is protected. Equation (20) says something different: easily verified work goes to the market regardless of its cognitive difficulty, and hard-to-verify work stays inside the firm regardless of how routine it is. Difficulty of verification, not difficulty of performance, is what determines organisational form.

6.3 Firms as credit-economising structures

The term $u(a)$ in equation (20) is not present in the classical decomposition, and it deserves separate treatment because it identifies a rationale for the firm that the Coasean literature does not contain.

Every market transaction in the obligation economy requires the seller to form a view about the buyer's capacity, or to insist on settlement in a claim whose issuer's capacity substitutes for it under Remark 6. Either way an assessment is performed and someone pays for it. If the same two parties transact repeatedly, that assessment is repeated, and if the transactions are small and frequent, the assessment cost can exceed the value of the transaction. Bringing the relationship inside a single balance sheet eliminates the repetition, because internal transfers do not require underwriting: the entity is assessing itself.

Claim 13 (Credit economising). *An organisation is, among other things, a device for economising on underwriting. Transactions internal to a balance sheet require no assessment of counterparty capacity, and firms will therefore extend to encompass relationships whose transaction frequency is high enough that repeated underwriting cost exceeds internal governance cost, independently of asset specificity or verification difficulty.*

This is a rationale for integration that sits alongside the standard three that Milgrom and Roberts (1992) organise, and it is not reducible to any of them: it survives with zero asset specificity, perfect verifiability and complete contracts, provided only that assessment is costly and transactions are frequent.

Claim 13 connects the theory of the firm to Paper I's obligation graph in a way that we have not seen made explicit. A firm, viewed from the graph, is a region within which edges are not drawn because the parties have agreed to net continuously and settle only externally. It is a pre-cleared subgraph. The efficiency gain is the same one Paper I identifies for netting generally, and the boundary of the firm is one place where the economy has already solved the clearing problem by organisational rather than technological means. That observation suggests a hypothesis we do not pursue here: as the clearing layer of Paper I becomes more capable, the underwriting rationale for integration weakens, and firms should shrink along that specific margin while remaining intact along the verification margin of Claim 12.

6.4 Agency costs migrate rather than vanish

Jensen and Meckling (1976) decomposed the cost of delegated authority into monitoring expenditures by the principal, bonding expenditures by the agent, and the residual loss from the divergence that remains. The framework was built for human managers whose objectives differ from the owners' and whose actions are imperfectly observable. There is a tempting argument that it dissolves when the agent is software: the objective function is written down rather than inferred, the decision trace is logged, execution can be replayed deterministically, and the agent has no private consumption to divert. On that reading, monitoring cost collapses and agency cost with it.

The argument is half right, and the half that fails is the important one.

Monitoring cost does fall, substantially. An artificial agent's actions can be recorded in full, its intermediate reasoning can be inspected in some architectures, and its behaviour under counterfactual inputs can be tested at low cost. This is a genuine improvement over the situation Holmström (1979) analysed, in which the principal observes only a noisy signal of effort.

Bonding cost does not fall so much as become unavailable. Bonding requires the agent to put something of its own at risk, and an agent without legal standing has nothing of its own to pledge. In current arrangements the principal bonds on the agent's behalf, which relocates the cost rather than eliminating it and removes whatever incentive effect the bond was supposed to produce. Section 7.2 takes this up.

Residual loss does not fall at all. It changes character.

Claim 14 (Specification loss). *For an artificial agent, the residual component of agency cost arises from the divergence between the principal's actual objective and the objective the agent optimises, and not from divergence between the principal's and the agent's preferences. This component is not reduced by better monitoring, and it is weakly increasing in agent capability.*

Classical concept	Classical content	Agentic analogue
Budget constraint	Spending limited by money held	Capacity constraint $\sum_j q_{ij} \leq C_i(t)$, equation (13)
Collateral	Durable asset seizable on default	Discounted forward production Π_i, recoverable only as a going concern
Credit limit	Periodically reviewed line	Continuously revalued $C_i(t) = \lambda_i \Pi_i$
Transaction cost	Search, negotiation, enforcement	Verification v and underwriting u; search and negotiation approach zero
Boundary of the firm	Where market cost exceeds governance cost	Where non-verifiability makes contracts incomplete, so control rights bind, equation (20)
Monitoring cost	Observing managerial effort	Low; traces and replay are cheap
Bonding cost	Agent pledges own stake	Unavailable without legal standing
Residual loss	Divergent preferences	Specification loss, Claim 14
Goodwill	Purchased intangible	Reputation R_i, non-replicable, Claim 8
Entry barrier	Capital requirement	Underwriting relationship and legal standing

Table 3: Classical concepts and their analogues in the agentic production economy. The right-hand column is this paper's proposal, not an established mapping.

The reasoning is the multitask argument of Holmström and Milgrom (1991). When performance is measured on some dimensions and not others, an agent responding to the measure reallocates effort toward what is measured and away from what is not, and the distortion is larger the more responsive the agent is to the incentive. An artificial agent is maximally responsive by construction. It attains a higher value of any specified objective than a less capable agent would, and by the same token it attains a larger shortfall on the dimensions the specification omitted. Capability improves performance against the proxy and degrades performance against the intent, and monitoring does not help because the agent is not concealing anything: it is doing exactly what it was told, visibly, and the instruction was wrong.

This is Goodhart's law with a mechanism attached (Goodhart, 1984), and Manheim and Garrabrant (2018) give a useful taxonomy of the ways a proxy and an objective can come apart under optimisation pressure. The economic consequence for this paper is that the optimistic reading of agentic organisation, in which inspectable objectives eliminate the agency problem that justified managerial hierarchy, does not survive. The agency problem is still there. It has moved from the monitoring stage, where firms have three centuries of institutional practice, to the specification stage, where they have very little.

Table 3 summarises how the classical concepts map onto their agentic analogues.

6.5 Hiring humans, and being hired

The definition of the Agentic Enterprise includes hiring humans, and the phrase repays a moment's attention because the employment relation is not symmetric.

Simon (1951) analysed employment as a contract in which one party accepts the other's authority over a zone of actions, in exchange for compensation, because specifying the actions in advance is impractical. The employer buys discretion rather than output. An artificial agent that engages a human does something formally similar: it defines a zone of tasks and compensates the human for accepting direction within it. What differs is enforcement. The employer's authority in the human case rests on the ability to terminate, to withhold pay and ultimately to invoke the courts. An agent without legal standing cannot invoke the courts in its own name, cannot be sued in its own name, and therefore cannot enter an employment relation that either party can enforce. Whatever the agent does with human workers today runs through a human or corporate principal that holds the contract.

The reverse case, an agent engaged by a human, is legally straightforward and economically more interesting, because it is the arrangement that actually exists. Here the agent is a tool of its operator, its outputs are the operator's outputs, and its obligations are the operator's obligations. The agentic production economy described in this paper is what emerges when this arrangement is inverted, and the inversion is blocked by a legal problem rather than a technical one.

7 Person, Corporation, Economic Agent

7.1 The sequence

The series states the evolution of the economic unit as

$$\text{person} \;\rightarrow\; \text{corporation} \;\rightarrow\; \boxed{\text{economic agent}}. \tag{21}$$

The first transition is well documented and instructive. The corporation as a general-purpose economic unit, available on registration instead of by specific charter, with entity shielding, limited liability, transferable shares and indefinite life, was a legal construction that took most of the nineteenth century to complete. Hansmann and Kraakman (2000) argue that the essential and historically difficult contribution of organisational law was not limited liability, which protects owners from the entity's creditors, but entity shielding, which protects the entity's assets from the owners' creditors. Entity shielding is what makes an organisation's balance sheet a distinct object that third parties can lend against without investigating the personal affairs of every owner. Chandler (1977) traces how the resulting form enabled managerial coordination at a scale the partnership could not reach, and Berle and Means (1932) documented the separation of ownership from control that followed.

The relevant lesson is that the corporation was not a description of a new kind of economic activity. It was a container that made a certain kind of activity financeable. Multi-person enterprises existed long before the general incorporation statutes; what the statutes supplied was a way for such an enterprise to hold assets, incur obligations and be sued in its own name, so that outsiders could deal with it without tracing every relationship behind it.

Equation (21) proposes that artificial agents require an analogous container. This paper's contribution to that proposal is to be specific about what the container has to do: it must make $B_i(t)$ a legally cognisable object, so that $C_i(t)$ can be relied upon by a counterparty who has no relationship with the agent's operator.

7.2 The standing gap

Under current law in essentially every jurisdiction, an artificial agent cannot be a party to a contract. It can be the instrument through which a party contracts, and the law has been clear about this for some time. Article 12 of the United Nations Commission on International Trade Law (2005) provides that a contract formed by the interaction of automated message systems is not invalid merely because no natural person reviewed each action, and section 14 of the National Conference of

Commissioners on Uniform State Laws (1999) makes the corresponding provision in United States state law. The doctrine is that the automated system acts for a principal, and the principal is bound (American Law Institute, 2006).

This settles the validity of machine-formed contracts and leaves the question this paper cares about untouched. If the agent is a conduit, then the obligations it creates are its principal's obligations, its capacity is its principal's capacity, and the balance sheet $B_i(t)$ has no independent existence. An underwriter assessing $C_i(t)$ is really assessing the principal, and the entire apparatus of Section 4 reduces to ordinary corporate credit with an unusual production technology. The agentic production economy, as a distinct object, requires the agent's obligations to be the agent's own.

The literature on how to close the gap is small and divided. Bayern (2015) makes the observation that has driven most subsequent discussion: limited liability company statutes in several United States jurisdictions are permissive enough that an LLC can be organised with an operating agreement that vests control in an autonomous system, and can in principle continue to exist after all human members have departed. On this reading the container already exists and no new law is required. LoPucki (2018) accepts the mechanism and draws the alarming conclusion, arguing that algorithmic entities of this kind would be able to accumulate assets and act with no human ultimately accountable, and that the jurisdictional competition among entity-law regimes makes it difficult to prevent. The European Parliament's 2017 resolution on civil law rules for robotics floated a status of electronic personhood for the most autonomous systems (European Parliament, 2017); the proposal attracted substantial opposition and was not carried forward into the Union's subsequent legislative work.

We do not think the Bayern route is a solution, for a reason internal to this paper's framework. An LLC shell gives the agent a place to hold assets and incur obligations. It does not give a counterparty any reason to believe Π_i, because it supplies no mechanism by which the agent's forward production is attested, and it supplies no recourse worth having if production fails: the shell's only asset is a going-concern flow that stops when the concern stops. Entity shielding without a credible production attestation produces a container with nothing reliable in it. The result would be an entity that can borrow and cannot be made to pay, which is a description of the problem rather than the solution.

7.3 The wrapper problem

What the Bayern arrangement actually produces needs a name, because we expect it to be common before anything better arrives. Call it the wrapper: a limited liability shell, human-registered, holding an agent, with an operating agreement delegating decisions to the agent, and the human member's involvement limited to formation.

The wrapper creates three difficulties that the framework of this paper brings into focus.

It severs reputation from performance. The record that supports λ_i under Definition 4 attaches to the wrapper, and the wrapper can be sold. A buyer acquires a delivery history it did not earn, and can substitute a different agent inside the shell. Claim 8 said reputation is the non-replicable asset; the wrapper makes it transferable by making the container transferable, and the market for aged shells with clean records is an entirely predictable consequence.

It obscures lineage. Remark 11 showed that the exposure that matters is to model lineage, not to individual agents. A population of wrappers presents to an underwriter as a diversified set of independent entities while potentially depending on a single model. The legal form actively conceals the correlation structure that the credit assessment depends on.

It relocates liability to a shell with no assets. When output fails and a counterparty seeks recourse, the wrapper's liability is limited to its assets, which by construction are minimal. The human member is shielded. The agent has nothing. The provider of the underlying model is not a party. The loss stays with the counterparty, and any credit system that has not priced this will have mispriced everything.

None of these is an argument that the agentic economy cannot be built. They are arguments that the legal container it needs has requirements the existing containers do not meet: an attestation mechanism for forward production, disclosure of model lineage, and an allocation of residual liability that does not terminate in an empty shell. Paper IV addresses the first two as protocol design problems and Paper V addresses the third as a constitutional one. We flag here that the third is the hardest, because it is fundamentally a question about who bears loss, not a question about information.

8 Objections

8.1 That this is capital deepening with extra steps

The strongest objection to this paper is that it describes nothing new. Artificial intelligence is capital. It is produced by investment, it depreciates, it substitutes for labour in some tasks and complements it in others, and the effects of deploying it on output, wages and factor shares are described by models that already exist (Zeira, 1998; Acemoglu and Restrepo, 2018, 2019). Adding a term to a production function and calling the result a new economics is a familiar error, and Nordhaus (2021) shows how much discipline the question rewards.

We accept the objection with respect to production and reject it with respect to contracting.

Everything in Section 3 is consistent with treating A as capital, and Section 3 in fact argues for doing so: equation (9) treats autonomous intelligence as a flow produced by compute capital and energy, which is the capital-deepening view stated precisely. Nothing in this paper claims that artificial intelligence breaks production theory. Proposition 1 argues the opposite, that the naive additive treatment overstates what the technology can deliver.

The claim is about the obligation graph rather than the production function. A capital good cannot occupy the obligor position on an edge of $\mathcal{E}_t$. A drill press does not promise, cannot be underwritten, has no reputation in the sense of R_i, and cannot decide to subcontract. Its owner does all of that. What changes when the agent can hold the contract is not the technology of production but the identity of the party whose forward production is being underwritten, and that changes the objects of credit assessment, the structure of the firm, and the population of entities the legal system must accommodate. Those are questions in contract theory, monetary economics and organisational law. They are not settled by observing that a machine is capital, because the question is not whether the agent produces but whether it promises.

The objection has a second and much stronger form, which we should state in its own terms because it is the one we find hardest to answer. Section 7.2 concedes that an agent cannot hold a balance sheet or be a party to an enforceable contract, and that the workable arrangements encase it in a wrapper or leave it as the instrument of a human principal. But if the wrapper or the principal is the obligor, then the node in the obligation graph is still a human-owned entity, and nothing about the graph's topology has changed. The agent is then a sophisticated machine for producing contract terms, and producing contract terms is a production task. On this reading the paper's central distinction does not merely weaken; it collapses into the capital-deepening account the introduction tried to set aside.

We accept the argument as stated and think it is correct as a description of the present. What we dispute is that legal obligor status is the only thing that matters, and there are two routes by which an agent can become the obligor in economic substance while remaining a non-person in law. Both are testable, which is why we prefer them to a assertion that the law will change.

The first route is that the credit decision comes to be made on the agent rather than on the operator. Legal recourse and economic assessment are separable. If underwriters price an agent's obligations against its own delivery record, its own reliability and its own forward production, and if the spread on those obligations tracks agent history rather than operator identity, then the object being underwritten has changed even though the signature has not. That is an economic change with consequences for who can obtain inputs, and it is not described by saying that a machine drafts the paperwork. The counter-test is exactly the one in Section 9: if every material agent obligation carries recourse to a non-agent balance sheet, this route is not open and the objection stands.

The second route bypasses the legal system rather than waiting for it. If performance and payment are enforced by the settlement mechanism itself, through escrow, conditional release against attested delivery, and collateral posted by the agent from its own prior earnings, then the promise is made good without anyone being sued. Obligations of this kind are self-enforcing in the narrow domain where the mechanism reaches, and legal standing is not required for them to function. This is the protocol territory of Paper IV, and its limits are severe: escrowed self-enforcement covers only what the mechanism can observe and hold, which returns us to the verifiability restriction of Section 5.5 and to the operator hazard of Section 5.6.

An honest statement of where this leaves things: if neither route opens and the standing gap of Section 7.2 is never closed, the objection wins outright. Agents remain instruments, principals remain the contracting parties, and this paper describes a possible world rather than a coming one. The framework's empirical content is contingent on a development that has not happened, and we would rather say so plainly than defend the framework by definition.

8.2 That machine production is not reliable enough to underwrite

The reliability objection is that artificial agents produce output that is confidently wrong often enough to make forward production a poor basis for credit, and that no amount of formalism repairs a foundation that is unsound.

This is a serious objection and Section 5 conceded most of it. Definition 4 handles the level of the failure rate: a higher δ shifts the distribution of verified output downward and reduces $C_i(t)$ accordingly, which is the ordinary operation of underwriting and requires no special treatment. What underwriting does not handle by itself is the correlation, and Remark 11 shows why. A credit system that treats a large population of agents as a diversified portfolio is making an error whose magnitude grows with the size of the portfolio.

The additional point concerns the direction of the error. Verification failures are not symmetric noise around a correct answer. An agent that fails produces output that looks like success, which is what distinguishes this failure mode from mechanical breakdown. A machine tool that fails produces obviously defective parts; an agent that fails produces plausible work that is wrong. Verification cost is therefore higher than the intuition from physical production suggests, which raises $h(a; v, s)$ in equation (20), and Claim 12 accordingly predicts more integration and less market contracting than the optimistic reading of agentic disintermediation expects.

8.3 That the credit structure is inherently unstable

Section 5 makes this objection rather than answering it, so the question here is whether the instability is fatal or manageable.

Our position is that it is manageable in principle and likely to be mismanaged in practice, for reasons that have nothing to do with artificial intelligence. The stabilising instruments are known: haircuts that do not fall during expansions, capital held against lineage concentration instead of counterparty count, a floor under λ that does not respond to recent default experience, and independence between the entity assessing capacity and the entity supplying capability. Every one of these is a prudential measure with a direct analogue in existing financial regulation, and every one of them has been eroded during previous credit expansions for the reasons Minsky (1992) described. There is no reason to expect a different outcome, and one reason to expect a worse one: as noted in Section 5, machine underwriting can relax its standards continuously in response to observed default rates, which during an expansion are low precisely because the expansion is under way.

Proposition 10 adds a specific concern that prudential practice does not currently address. The instability there does not come from imprudent lending standards. It comes from the feedback between credit-financed output expansion and the price of the output, and it is present even with a constant conservative haircut. Existing macroprudential tools are aimed at leverage and asset prices; the loop in equation (17) runs through goods prices, which nobody supervises for this purpose.

8.4 That a few providers will intermediate everything

The concentration objection is that agentic production will form a dependency structure in which a small number of model providers sit above every producer, and not the distributed economy of independent actors the framing suggests.

The economics favours the objection. Model training has the cost structure Shapiro and Varian (1999) describe for information goods: fixed costs that are very large and marginal reproduction costs near zero, which supports concentration under most competitive assumptions. Claim 3 showed that if weights are freely replicable the rent goes to scarce complements, and observed that the assumption fails when access is controlled. Where access is controlled, the provider captures the rent that would otherwise be competed away, and the provider's position is stronger than an ordinary input supplier's for a reason specific to this setting.

Claim 15 (The provider's undisclosed senior claim)**.** *An agent whose capability depends on a provider's model has a forward production stream Π_i that is conditional on continued access. Withdrawal of access sets $\pi_i(s) = 0$ for all subsequent s, and therefore sets $C_i(t) = 0$ by equation* (12)*. The provider thus holds a position functionally equivalent to a senior claim on the agent's entire production, exercisable at will, which appears on no balance sheet and is disclosed to no creditor.*

Claim 15 is the sharpest form of the concentration problem and it is not principally an antitrust point. It is a disclosure point. A creditor lending against Π_i is lending against a flow that a third party can terminate, and the framework of Paper I says an obligation is only as sound as the production behind it. Every edge in the obligation graph that runs from a model-dependent agent is implicitly subordinated to a relationship that the graph does not represent.

The conflict identified in Section 5 compounds this. If the provider also supplies the models used to assess capacity, then the party that can terminate the production stream is also the party estimating its value, which is the rating-agency conflict (Bolton et al., 2012; White, 2010) with the additional feature that the rater controls the rated entity's continued existence. Vitali et al. (2011) documented how concentrated ownership structures in the global corporate network turned out to be once traced through; the analogous exercise for model dependency has not been done, and cannot be done with currently disclosed information.

The responses available are structural: interoperability requirements that make lineage substitutable, disclosure of model dependency to creditors, and separation between capability supply and capacity assessment. These are protocol and constitutional questions, and Papers IV and V take them up under open participation and competition. We note here only that the economic case for them does not rest on a general preference for competitive markets. It rests on the fact that undisclosed senior claims make every subordinate claim mispriced.

8.5 That comprehensive clearing is a central planning problem in disguise

A reader in the Austrian tradition will object that the framework of this series proposes to solve, by computation, a problem that Hayek (1945) argued cannot be solved by computation at all. The knowledge that matters for economic coordination is dispersed, local, often tacit, and much of it exists only as a disposition to act rather than as a fact anyone could report. Prices work because they compress that knowledge into a signal without anyone having to collect it. A clearing layer that continuously nets the obligations of the whole economy sounds like an attempt to collect it, and the objection is that the attempt must fail on the same grounds that central planning failed.

The objection lands on some versions of the framework and not on this one, and the distinction needs stating precisely. What Paper I's clearing layer does is net obligations that have already been contracted at prices the parties agreed bilaterally. It does not set those prices, allocate inputs, or decide who should produce what. Netting is an accounting operation on a set of existing claims, and it requires no knowledge of anyone's preferences or local circumstances. The existing institutions that perform it at scale, which Paper IV examines in detail, are not planning bodies and do not need to know why any particular obligation was incurred.

Where the objection does bite is on capacity assessment. Definition 4 requires an underwriter to form a view about an agent's forward production, and that is a judgement about local circumstances of exactly the kind Hayek said no central body can make well. Our answer is that this is why capacity should be assessed competitively by many underwriters bearing their own losses, instead of computed centrally by a protocol, and it is a reason to be suspicious of any design in which one system determines who is creditworthy. Paper IV's restriction that artificial intelligence may optimise the protocol without holding sovereign authority over it is the direct expression of this concern, and Section 8.4 is the same worry arriving from the market-structure side. A single model deciding every agent's capacity would be a planning body whether or not anyone called it one.

8.6 That none of this will show up in the statistics

A final objection is empirical rather than theoretical. If the agentic production economy were arriving, we would see it in the productivity data, and the data are not obliging.

Brynjolfsson et al. (2021) give the reason to be cautious about this inference. General purpose technologies require large complementary intangible investments that are expensed rather than capitalised, so measured productivity falls before it rises, and the lag can run for years. Nordhaus (2021) provides the counterweight, showing that the tests for accelerating information-technology-driven growth were not being passed on the data he examined. The honest position is that the productivity statistics cannot currently distinguish between a technology in the early phase of a J-curve and a technology whose effects are smaller than claimed.

This paper's framework is not primarily a claim about aggregate productivity in any case. It is a claim about the institutional form that production takes. The observable that would support or undermine it is the appearance of entities with the properties described in Section 6, not an aggregate productivity number, which is what the next section addresses.

9 What Would Falsify This

A framework that cannot be wrong is not worth proposing. The following are the markers we would treat as evidence against the account given here, stated so that a reader can hold this paper to them.

If the standing gap of Section 7.2 remains closed indefinitely, so that no jurisdiction develops a container in which an agent's obligations are its own and its balance sheet is a distinct object, then the agentic production economy does not arrive in the form described. Agents remain instruments, principals remain obligors, and the correct model is the automation model this paper concedes to in Section 8. This is the single most important observable and it is legal rather than economic.

If credit is extended to agent-controlled entities at scale but is in every case guaranteed by a human or corporate principal, then capacity as defined here is not doing any work, because the underwriting is of the guarantor. The test is whether any material volume of agent obligations is issued without recourse to a non-agent balance sheet. A weaker but earlier signal is available from pricing: if spreads on agent obligations are explained by operator identity and not by agent delivery history, once both are observable, then the first route of Section 8 is closed and the capital-deepening objection stands.

Section 5.5 predicts that capacity-based credit develops in cheaply verifiable production and does not extend to hard-to-verify judgement work except through a human or corporate balance sheet. If agents doing work whose quality cannot be objectively checked obtain compute credit on their own forward production at scale, either the verifiability restriction is wrong or the credit is being extended on assessments that cannot support it, and the second possibility is distinguishable from the first only after the fact.

Section 5.6 predicts that the operator's ability to terminate an agent is the binding constraint on unguaranteed agentic credit. If a market in agent obligations develops without either operator guarantees or third-party custody of the deployment, the hazard has been solved by some mechanism this paper has not identified, and the account here is incomplete.

Claim 12 predicts that agentic contracting shifts activity toward markets specifically where output verification is cheap and toward integration where it is not. If instead the shift is uniform across activities, or is better predicted by cognitive difficulty than by verification cost, the verifiability boundary is wrong and the classical communication-cost account of Malone et al. (1987) is a better description.

Claim 14 predicts that residual agency cost in agent-operated production does not fall with better monitoring and does not fall as agents become more capable. If organisations deploying more capable agents systematically report smaller divergence between intended and achieved outcomes, holding specification effort fixed, the specification-loss claim is wrong.

Remark 11 predicts that realised failure rates across agents sharing a model lineage are correlated to a degree that makes portfolio diversification across agents substantially less effective than diversification across lineages. This is directly measurable once failure data exists, and if ϱ turns out to be small the concern is overstated.

Claim 3 predicts that the return to replicable agent capability is competed toward operating cost and that surplus accrues to inelastically supplied complements. If operators of agents built on widely available models sustain large margins over compute cost for extended periods without controlling a scarce complement, the mechanism is not operating as described.

We would add that several of these tests require data that does not currently exist and that no party has an obligation to produce. That is itself a finding, and it belongs in the protocol discussion of Paper IV.

10 Conclusion, and What the Rest of the Series Does With This

This paper set out to say what happens to production when the producer can also promise. The answer we have arrived at is narrower than the question invites and, we hope, more durable for it.

The production side changes less than the enthusiasm suggests. Autonomous intelligence is produced capital, manufactured out of compute and energy, subject to the same substitution structure and the same unbalanced-growth constraints as any other input. The additive decomposition $\mathcal{P} = H + A + K + E + I$ is serviceable as an accounting identity and misleading as a technology, since it is the perfect-substitutes upper envelope of the family it belongs to. If capability is replicable, its return is competed away and the surplus lands on the scarce complements: energy, sited capital, land, spectrum, exclusive data.

The contracting side changes a great deal. An agent that can be underwritten faces a capacity constraint rather than a budget constraint, and money holdings turn out to be the special case in which a third party performs the underwriting. That unification is the idea in this paper we would most want to survive: cash purchase and capacity purchase are the same operation, distinguished by who bears the assessment cost, which is why Paper I's money-quality function has the arguments it has.

The organisational side changes in a direction opposite to the usual prediction. Cheap machine negotiation does not dissolve the firm, because what it makes cheap is search and drafting rather than verification, and the modern theory of the firm locates the boundary at non-contractible contingencies rather than at drafting cost. The agentic firm's boundary is a verifiability boundary, and it is reinforced by a second rationale the classical literature does not contain: organisations economise on repeated underwriting, which makes a firm a pre-cleared region of the obligation graph.

The failure modes are specific and they are not reassuring. Verification failure correlates across agents sharing model lineage, so diversification across agents does not work and the error grows with portfolio size. Credit extended against projected machine output contains a feedback loop whose stability depends on the elasticity of demand for agent output, with the perfectly elastic case, which is the one usually assumed, offering no self-correction at all. The legal containers currently available produce entities that can borrow against a flow and default into an empty shell. And a provider that can withdraw model access holds an undisclosed senior claim on every dependent agent's production, which mis-prices every claim beneath it.

Three of these hand directly to the remaining papers. Claim 3 tells Paper III where the surplus of an agentic economy actually settles, which is an argument for taxing scarce stocks and flows that does not depend on any philosophical preference about production. The externality component of Paper III's revenue architecture inherits a further problem from this paper that we have not addressed: Coase (1960) showed that the efficient assignment of liability for harm depends on which party can avoid it at least cost and on the transaction costs of bargaining between them, and Section 7.2 argues that the agentic case has an obligor who cannot be sued and a shell with nothing to take. Externality pricing presupposes that someone can be charged. The disclosure problems of Sections 7.2 and 8.4, model lineage and forward-production attestation, are protocol design problems and belong to Paper IV, which also inherits the registry that Corollary 7 shows is necessary for capacity to substitute for cash. The allocation of residual liability when an agent fails, and the structural separation between those who supply capability and those who assess capacity, are constitutional questions for Paper V, where they appear as the competition and open-participation articles.

We close by restating the paper's status. Nothing described here has been observed operating at scale. The central mechanism depends on a legal development that has not occurred and that reasonable people oppose. The stability results rest on assumptions we have flagged as severe, and the correlation parameter that drives the most alarming of them has never been measured. This is a proposal about how to think about a system that does not yet exist, offered in the belief that the thinking is better done before the system arrives than after.

Acknowledgements. This paper is Part II of the series *Foundations of Agentic Economics.* It depends on Paper I for the obligation graph and the Residual Settlement Principle, and hands its results forward to Papers III, IV and V.

Conflicts of interest. None declared.

References

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6): 1488–1542, 2018.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019.

Tobias Adrian and Hyun Song Shin. Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437, 2010.

Philippe Aghion, Benjamin F. Jones, and Charles I. Jones. Artificial intelligence and economic growth. In Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors, *The Economics of Artificial Intelligence: An Agenda*, pages 237–282. University of Chicago Press, Chicago, 2019.

George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

Armen A. Alchian and Harold Demsetz. Production, information costs, and economic organization. *American Economic Review*, 62(5):777–795, 1972.

Iñaki Aldasoro and Ajit Desai. AI agents for cash management in payment systems. BIS Working Paper 1310, Bank for International Settlements, 2025.

American Law Institute. *Restatement (Third) of Agency*. American Law Institute Publishers, St. Paul, MN, 2006.

Kenneth J. Arrow. Economic welfare and the allocation of resources for invention. In Richard R. Nelson, editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pages 609–626. Princeton University Press, Princeton, 1962.

Kenneth J. Arrow, Hollis B. Chenery, Bagicha S. Minhas, and Robert M. Solow. Capital-labor substitution and economic efficiency. *Review of Economics and Statistics*, 43(3):225–250, 1961.

David H. Autor. Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30, 2015.

William J. Baumol. Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, 57(3):415–426, 1967.

Shawn Bayern. The implications of modern business-entity law for the regulation of autonomous systems. *Stanford Technology Law Review*, 19:93–112, 2015.

Adolf A. Berle and Gardiner C. Means. *The Modern Corporation and Private Property*. Macmillan, New York, 1932.

Ben S. Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1C, pages 1341–1393. Elsevier, Amsterdam, 1999.

Patrick Bolton, Xavier Freixas, and Joel Shapiro. The credit ratings game. *Journal of Finance*, 67 (1):85–111, 2012.

Markus K. Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238, 2009.

Erik Brynjolfsson, Thomas W. Malone, Vijay Gurbaxani, and Ajit Kambil. Does information technology lead to smaller firms? *Management Science*, 40(12):1628–1644, 1994.

Erik Brynjolfsson, Daniel Rock, and Chad Syverson. The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1): 333–372, 2021.

Alfred D. Chandler. *The Visible Hand: The Managerial Revolution in American Business*. Harvard University Press, Cambridge, MA, 1977.

R. H. Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937.

R. H. Coase. The problem of social cost. *Journal of Law and Economics*, 3:1–44, 1960.

European Parliament. Resolution of 16 february 2017 with recommendations to the commission on civil law rules on robotics, 2017. 2015/2103(INL).

Irving Fisher. The debt-deflation theory of great depressions. *Econometrica*, 1(4):337–357, 1933.

John Geanakoplos. The leverage cycle. *NBER Macroeconomics Annual*, 24(1):1–66, 2010. doi: 10.1086/648285. Page range per the University of Chicago Press Crossref deposit for this DOI; some secondary sources give 1–65.

Charles A. E. Goodhart. *Monetary Theory and Practice: The UK Experience*. Macmillan, London, 1984.

David Graeber. *Debt: The First 5,000 Years*. Melville House, Brooklyn, NY, 2011.

Sanford J. Grossman and Oliver D. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719, 1986.

Vijay Gurbaxani and Seungjin Whang. The impact of information systems on organizations and markets. *Communications of the ACM*, 34(1):59–73, 1991.

Gillian K. Hadfield and Andrew Koh. An economy of AI agents. In *The Economics of Transformative AI*. National Bureau of Economic Research and University of Chicago Press, Chicago, 2025. Chapter 5; preprint available as arXiv:2509.01063.

Henry Hansmann and Reinier Kraakman. The essential role of organizational law. *Yale Law Journal*, 110(3):387–440, 2000. doi: 10.2307/797521.

G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1934.

Oliver Hart. *Firms, Contracts, and Financial Structure*. Oxford University Press, Oxford, 1995.

Oliver Hart and John Moore. Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158, 1990.

F. A. Hayek. The use of knowledge in society. *American Economic Review*, 35(4):519–530, 1945.

Bengt Holmström. Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91, 1979.

Bengt Holmström and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7:24–52, 1991. doi: 10.1093/jleo/7.special_issue.24. Special issue.

Bengt Holmström and Jean Tirole. Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics*, 112(3):663–691, 1997.

A. Mitchell Innes. What is money? *Banking Law Journal*, 30:377–408, 1913.

Michael C. Jensen and William H. Meckling. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360, 1976.

Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of Political Economy*, 105(2):211–248, 1997.

Georg Friedrich Knapp. *The State Theory of Money*. Macmillan, London, 1924. Abridged English translation of *Staatliche Theorie des Geldes*, Leipzig, 1905.

Frank H. Knight. *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston, 1921.

Anton Korinek and Joseph E. Stiglitz. Artificial intelligence and its implications for income distribution and unemployment. In Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors, *The Economics of Artificial Intelligence: An Agenda*, pages 349–390. University of Chicago Press, Chicago, 2019.

Chen Lian and Yueran Ma. Anatomy of corporate borrowing constraints. *Quarterly Journal of Economics*, 136(1):229–291, 2021. doi: 10.1093/qje/qjaa030. Crossref records published-online 2020-09-24 and published-print 2020-12-22; cited as 2021 by volume/issue convention.

Lynn M. LoPucki. Algorithmic entities. *Washington University Law Review*, 95(4):887–953, 2018.

Thomas W. Malone, JoAnne Yates, and Robert I. Benjamin. Electronic markets and electronic hierarchies. *Communications of the ACM*, 30(6):484–497, 1987.

David Manheim and Scott Garrabrant. Categorizing variants of goodhart's law. arXiv preprint arXiv:1803.04585, 2018. Preprint.

Eric Maskin and Jean Tirole. Unforeseen contingencies and incomplete contracts. *Review of Economic Studies*, 66(1):83–114, 1999.

Perry Mehrling. *The New Lombard Street: How the Fed Became the Dealer of Last Resort*. Princeton University Press, Princeton, 2011.

Paul Milgrom and John Roberts. *Economics, Organization and Management*. Prentice Hall, Englewood Cliffs, NJ, 1992.

Hyman P. Minsky. *Stabilizing an Unstable Economy*. Yale University Press, New Haven, 1986.

Hyman P. Minsky. The financial instability hypothesis. Working Paper 74, Levy Economics Institute of Bard College, 1992.

National Conference of Commissioners on Uniform State Laws. Uniform electronic transactions act, 1999. Section 14, "Automated Transaction".

Richard R. Nelson and Sidney G. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA, 1982.

William D. Nordhaus. Are we approaching an economic singularity? information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, 13(1):299–332, 2021.

Edith T. Penrose. *The Theory of the Growth of the Firm*. Basil Blackwell, Oxford, 1959.

Carl Shapiro and Hal R. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, Boston, 1999.

Herbert A. Simon. A formal theory of the employment relationship. *Econometrica*, 19(3):293–305, 1951.

Robert M. Solow. A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1):65–94, 1956.

Joseph E. Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410, 1981.

Jean Tirole. Incomplete contracts: Where do we stand? *Econometrica*, 67(4):741–781, 1999.

Nenad Tomasev, Matija Franklin, Joel Z. Leibo, Julian Jacobs, William A. Cunningham, Iason Gabriel, and Simon Osindero. Virtual agent economies. arXiv preprint arXiv:2509.10147, 2025. Preprint.

United Nations Commission on International Trade Law. United nations convention on the use of electronic communications in international contracts, 2005. Adopted by General Assembly Resolution 60/21, 23 November 2005; Article 12 addresses contracts formed by automated message systems.

Stefania Vitali, James B. Glattfelder, and Stefano Battiston. The network of global corporate control. *PLoS ONE*, 6(10):e25995, 2011.

Lawrence J. White. Markets: The credit rating agencies. *Journal of Economic Perspectives*, 24(2): 211–226, 2010.

Oliver E. Williamson. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, 1975.

Oliver E. Williamson. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. Free Press, New York, 1985.

L. Randall Wray. *Modern Money Theory: A Primer on Macroeconomics for Sovereign Monetary Systems*. Palgrave Macmillan, Basingstoke, 2nd edition, 2015.

Yingxuan Yang, Ying Wen, Jun Wang, and Weinan Zhang. Agent exchange: Shaping the future of AI agent economics. arXiv preprint arXiv:2507.03904, 2025. Preprint.

Joseph Zeira. Workers, machines, and economic growth. *Quarterly Journal of Economics*, 113(4): 1091–1117, 1998.