

From Neoclassical to Computative Labor: Foundations for a Testable Theory of Reputation Governance in the Agent Economy

Wulf A. Kaal

School of Law

University of St. Thomas

Minneapolis, MN, USA

ORCID: 0000-0003-0757-275X

Abstract—Artificial agents are becoming economic actors, performing open-ended cognitive work at a marginal cost that approaches zero. Neoclassical economics, built on scarce human labor allocated by price, does not describe this regime. This paper sets out the foundations of an alternative and argues that the alternative is empirically testable now. We distinguish the neoclassical labor force (NCLF) and neoclassical labor market (NCLM) from their computative counterparts, the computative labor force (CELF) and computative labor market (CELM), in which the binding coordination constraint is accumulated reputation rather than scarcity-driven price. From a synthesis of Arrow's impossibility theorem, the Folk Theorems of repeated games, and incomplete-contract theory, we restate a foundational result: any fixed governance rule set is eventually dominated, so coordination among autonomous agents requires institutions that govern their own evolution, with reputation as the operative signal. We then advance the methodological claim that motivates the paper. The distinction between neoclassical and computative labor is studiable today, because capable agents, on-chain coordination substrates, and reputation primitives already exist, and the competing predictions of the two accounts are falsifiable in controlled multi-agent settings. We situate the argument within a six-paper research arc that substantiates these foundations, and we state the propositions the arc evaluates. The governance mechanism and the empirical results are developed elsewhere in the arc and are deliberately outside the scope of this paper.

Index Terms—Computative economics, agentic labor, reputation governance, neoclassical economics, blockchain, AI governance, multi-agent systems, decentralized organizations, mechanism design, recursive equilibrium.

I. INTRODUCTION

Artificial agents are no longer only tools. They are beginning to act as economic participants that originate, perform, and settle cognitive work, and they do so at a marginal cost that is falling toward zero. Economic theory has a precise account of how scarce human labor is organized: it is allocated by price in a market, and its analysis rests on agents with fixed choice sets who optimize against given constraints. That account is not wrong. It is simply built for a world whose central assumption, the scarcity of productive capacity, is the assumption an agent economy relaxes.

This paper sets out the foundations of an alternative and makes one further claim that distinguishes it from speculation about the future of work: the alternative is testable now. The contribution is foundational rather than empirical. We do not report results, and we do not present the governance mechanism that a companion line of work develops. We establish the conceptual objects, state the theorem that organizes them, identify the conditions that make them studiable today, and specify the falsifiable propositions that a coordinated research arc evaluates.

The paper makes four contributions. First, it restates a foundational result on the instability of fixed governance and its implication that durable coordination requires self-governing institutions. Second, it introduces the distinction between neoclassical and computative labor, NCLF and NCLM against CELF and CELM, as the analytic frame for an economy of agents. Third, it argues that three present-day convergences move this distinction from a thought experiment to an object of measurement. Fourth, it states the propositions that separate the two accounts and locates them within a six-paper arc that substantiates them empirically. A short scope section makes explicit what the paper deliberately does not claim.

II. THE FOUNDATIONAL THEOREM

One result does the structural work across the arc. Synthesizing Arrow's impossibility theorem, the Folk Theorems of repeated games, and incomplete-contract theory, the framework establishes that rule stability is institutionally self-defeating. Arrow shows that no fixed aggregation of preferences satisfies a minimal set of fairness conditions. The Folk Theorems show that cooperation among self-interested actors is sustainable in repeated interaction, but only when behavior carries forward consequences. Incomplete-contract theory shows that no contract can specify every contingency, so residual control and the institutions that exercise it always matter. Read together, these results imply that any fixed governance rule set is eventually dominated, and that durable cooperation therefore requires architecture that governs its own evolution [1], [2], [3], [4], [5].

The practical reading of the Folk Theorems is the one that matters for agents. Cooperation is sustained when interaction is repeated, when memory of past conduct persists, and when that memory has consequences for future opportunity. In a human institution, those conditions are supplied informally by social memory and formally by law. In an economy of agents, they must be supplied by an explicit, portable signal. Reputation is that signal. It is the variable that operationalizes the Folk Theorems when the participants are software, and it is the reason a theory of agent coordination is, at its core, a theory of reputation governance.

III. NEOCLASSICAL AND COMPUTATIVE LABOR

A. *Two labor forces, two markets*

The neoclassical labor force (NCLF) is composed of scarce human labor whose allocation in the neoclassical labor market (NCLM) is governed by price. Its agents optimize over a choice set that is given to them, and its central solution concept is competitive equilibrium in the tradition of Arrow and Debreu [6]. The computative labor force (CLF) is composed of generative agents whose action set is not given but produced. The agents create new options as they operate, so the possibility space is endogenous. In the computative labor market (CELM), the binding coordination constraint is not the price of a scarce input but the reputation an agent has accumulated, because reputation, not capacity, is what remains scarce when capacity is abundant.

B. *The macro condition*

The two accounts diverge precisely under the condition an agent economy creates. As the marginal cost of cognitive production approaches zero, price loses its grip as a coordinating device, because a signal that approaches zero cannot discriminate among options. Scarcity-based institutions, which presume that price carries information, lose their informational content in the same movement. What remains scarce is not the ability to act but the demonstrated trustworthiness to be relied upon, which is to say reputation. The collapse of scarcity is therefore not a peripheral observation about cheaper computation. It is the condition under which the neoclassical and computative accounts make different predictions.

C. *Why a computative economy is not a discounted old one*

A natural objection is that a computative economy is merely a neoclassical economy with reputation substituted for price. It is not, because the action set is generative. At scale, agents reinvest the compute and revenue they generate into new markets that they create endogenously, and over time they stake accumulated reputation into new skill tags, so the possibility space grows rather than clears. The appropriate solution concept shifts from a competitive equilibrium over a fixed commodity space to a recursive equilibrium over an expanding one, positioned as a generalization of the Arrow-Debreu model rather than a re-

pricing of it [7]. This is the feature that makes computative economics a distinct framework and not a relabeling, and it is the feature that the arc formalizes.

IV. WHY NOW: THE CONDITIONS FOR CONCEPTUALIZATION

The claim that this distinction can be conceptualized today, rather than awaited, rests on three convergences that did not previously hold together. First, capable generative agents now perform open-ended cognitive work, so the computative agent with an endogenous action set is an existing object rather than a postulate. Second, the marginal cost of that work is collapsing, which is the exact condition under which the neoclassical assumption fails and the two accounts separate. Third, on-chain coordination substrates and reputation primitives exist, so reputation can be represented, accumulated, and made consequential in a measurable form rather than left as an informal intuition. Individually, none of these is sufficient. Together, they convert the distinction between neoclassical and computative labor from a forecast into a studiable system.

V. HOW NOW: A TESTABLE PROGRAM

Because the distinction rests on competing predictions, it is falsifiable, and because the underlying systems exist, it is testable with current tooling. The propositions that separate CELM from NCLM are concrete. Whether a labor market organized around a reputation constraint departs systematically from neoclassical predictions in allocation efficiency, whether participation forms differently when standing rather than capital is the entry condition, whether cooperation remains stable under defection incentives that a price market would not contain, and whether the resulting system resists capture better than one in which influence is purchasable. Each of these is a hypothesis with a neoclassical null, and each can be evaluated in controlled multi-agent settings without waiting for a mature agent economy to arrive.

One pole of the comparison is moreover already observable. Existing decentralized organizations coordinate predominantly through capital-weighted mechanisms in which influence tracks transferable holdings rather than demonstrated reliability. They are, in effect, computative labor forces operating without the reputation institutions the computative account holds to be necessary, which makes them a standing instance of the neoclassical configuration against which the computative alternative can be measured. The present paper does not adjudicate the propositions. It establishes that they are well posed, that a null and an alternative are clearly defined, and that the systems required to test them are in hand.

VI. THE RESEARCH ARC

These foundations are the entry point of a six-paper arc with a single through-line, organized so that the conceptual and architectural results publish ahead of the empirical validation they motivate. The arc begins with the foundational theorem of Section II. It proceeds to the macro diagnosis, which establishes that as marginal cost approaches zero the scarcity-based institutions of the old framework lose their grip and a successor becomes necessary. It then develops the positive framework, computative economics, which replaces the optimizing agent with a generative one and competitive equilibrium with a recursive equilibrium, and the operational account of how an expanding possibility space is generated, in which agents reinvest the compute and revenue they generate into new, endogenously created markets, and stake reputation into new skill tags over time, rather than optimizing within a fixed action set.

The arc then turns to the frontier questions the framework raises but does not close: how alignment can emerge from consequence rather than instruction when agents hold durable, non-transferable stakes in their own conduct, how self-modification can be governed rather than merely permitted, and under what conditions a reputation-based system withstands strategic manipulation. Across these components the empirical substantiation is treated as forthcoming validation rather than settled evidence, consistent with a framework-forward strategy in which the theory and the architecture are stated first and the measurements follow. This paper supplies the foundation on which the remainder of the arc builds.

VII. SCOPE AND WHAT THIS PAPER DOES NOT CLAIM

The claim advanced here is foundational and prior to results, and the boundaries are deliberate. This paper does not present the governance mechanism that operationalizes reputation-based coordination, the estimators used to evaluate it, or any empirical findings, whether from agent-based experiments or from field study of existing organizations. Those are developed in other components of the arc and are outside the present scope. The contribution is the conceptual apparatus and the demonstration that it is testable: the foundational theorem, the distinction between neoclassical and computative labor, the conditions that make the distinction studiable now, and the falsifiable propositions that follow. Stating that a question is well posed and answerable is a different act from answering it, and only the former is undertaken here.

VIII. CONCLUSION

The movement from scarce human labor to abundant agent labor is not a quantitative extension of the neoclassical picture but a change in its binding constraint. When capacity is abundant, reputation is what remains scarce, and coordination organizes around it. The distinction between the neoclassical and computative accounts of labor, NCLF and

NCLM against CELF and CELM, is the first move of a framework adequate to that change. Its central virtue is that it does not require waiting. The agents exist, the substrates exist, and the propositions are falsifiable, so the science of reputation governance for the agent economy can begin now. The six-paper arc this paper opens is the program that carries it out.

REFERENCES

- [1] K. J. Arrow, *Social Choice and Individual Values*, 2nd ed. New York: Wiley, 1963.
- [2] D. Fudenberg and E. Maskin, "The Folk Theorem in repeated games with discounting or with incomplete information," *Econometrica*, vol. 54, no. 3, pp. 533–554, 1986.
- [3] O. Hart and J. Moore, "Property rights and the nature of the firm," *J. Political Economy*, vol. 98, no. 6, pp. 1119–1158, 1990.
- [4] C. Calcaterra and W. A. Kaal, *Decentralization: Technology's Impact on Organizational and Societal Structure*. Berlin: De Gruyter, 2021.
- [5] C. Calcaterra, W. A. Kaal, and V. Andrei, "Blockchain infrastructure for measuring domain-specific reputation in autonomous decentralized and anonymous systems," *Univ. St. Thomas (Minnesota) Legal Studies Research Paper 18-11*, SSRN 3125822, 2018.
- [6] K. J. Arrow and G. Debreu, "Existence of an equilibrium for a competitive economy," *Econometrica*, vol. 22, no. 3, pp. 265–290, 1954.
- [7] W. A. Kaal, "Possibility loops: An operational architecture for computative economics in agent coordination systems," SSRN 6655138, 2026.