Information Frictions and Innovation: A Formal Theory

[o3 Deep Research] *

February 2, 2025

Abstract

This paper was written with a one-shot prompt (from Kevin Bryan) on o3 Deep Research, no iteration, 10 minutes of thinking. This paper develops a formal economic theory exploring how *information frictions* impact innovation, extending beyond the usual focus on incentive problems. We present a model of innovation in which the production of new ideas builds on previous innovations, but knowledge about these prior innovations is distributed across many agents. In this environment, classical welfare theorems break down: key inputs into innovation (knowledge) are unpriced and information is not optimally aggregated, leading to market failures. We formally compare several mechanisms — patents, prizes, advance market commitments (AMCs), and others — in their ability to overcome these information frictions. We derive propositions showing how each mechanism influences the aggregation of dispersed knowledge and the efficiency of innovation, providing rigorous proofs. Our results highlight that beyond providing incentives, innovation institutions serve a critical role in coordinating distributed information. The analysis yields insights into the design of innovation policy when knowledge is decentralized.

1 Introduction

Innovation is widely recognized as a key driver of economic growth and social welfare. Traditional economic analyses of innovation focus on incentive problems: because new knowledge is non-rival and often non-excludable, markets left to themselves may under-provide innovative effort. Canonical work by Arrow (1962) formalized the idea that knowledge has public good characteristics, leading to underinvestment in R&D. Standard policy responses, such as patents or prizes, aim to correct this underinvestment by improving innovators' rewards. However, beyond underinvestment lies another fundamental challenge: *information frictions*. Knowledge required for innovation is often dispersed among many agents, and no single actor or market automatically aggregates this distributed information. These frictions can cause misallocation of innovative effort and failures of coordination that are not addressed by incentive alignment alone.

This paper develops a formal theory of innovation under distributed information. We ask: How do information frictions — such as dispersed private knowledge of research opportunities

*[OpenAI]. Correspondence: A Robot..

or the inability to fully communicate existing know-how — impede efficient innovation, even if incentive mechanisms are in place? We incorporate the idea that each innovation builds on prior knowledge, and that knowledge is spread across different people and firms. In this setting, we show that the classical welfare theorems fail: a decentralized market equilibrium will generally not achieve an optimal outcome, and one cannot simply assign property rights or lump-sum transfers to restore efficiency because key inputs are information that is not fully tradable or priced.

We then formally analyze and compare alternative mechanisms for encouraging innovation when information is dispersed. Patents, prizes, and advance market commitments (AMCs) are prominent policy tools; each provides rewards for innovation, but they differ in how they mobilize and aggregate knowledge. Patents rely on decentralizing decisions to innovators and create markets for technology through intellectual property rights. Prizes and AMCs involve a principal (like a government) who commits to reward outcomes, potentially guiding decentralized agents towards particular goals. We develop formal propositions to compare these mechanisms in terms of efficiency and information aggregation. Notably, we highlight that mechanisms which perform well are those that not only incentivize effort but also effectively coordinate the use of dispersed knowledge.

The contributions of the paper are both positive and normative. Positively, we extend the theory of innovation to explicitly incorporate distributed information, thereby generalizing the classic Arrow (1962) framework which focused on appropriation incentives. Normatively, we identify conditions under which one innovation policy dominates another. For example, we show that when knowledge is highly fragmented across agents, a pure patent system may fail to achieve many innovations due to coordination failures, whereas appropriately designed prize systems or collaborative mechanisms can better assemble the pieces of knowledge. We also formalize why the fundamental theorems of welfare economics break down in this context: the presence of unpriced knowledge externalities and incomplete information prevents markets from achieving efficiency.

The rest of the paper is organized as follows. Section 2 reviews the related literature, including the foundational insights of Arrow and Hurwicz on innovation and mechanism design, and more recent works on innovation externalities and information distribution. Section 3 presents the model of innovation with distributed information, describing the technology, information structure, and the notion of equilibrium. Section 4 analyzes the inefficiencies in the decentralized outcome and formally demonstrates the failure of the welfare theorems. Section 5 introduces various innovation mechanisms (patents, prizes, AMCs, etc.) and provides formal propositions comparing their performance in our model, with proofs highlighting the role of information aggregation. Section 6 discusses the implications of the results and how hybrid or improved mechanisms might address both incentive and information problems. Section 7 concludes.

2 Literature Review

Our work builds on several strands of literature in economics: the economics of innovation and R&D incentives, the theory of mechanism design and information in markets, and studies of cumulative knowledge and innovation externalities. We briefly review the most relevant contributions.

Innovation as a public good and market failure. The idea that markets underprovide innovation due to the public good nature of knowledge dates back at least to Nelson (1959) and Arrow (1962). Arrow's seminal 1962 paper articulated the fundamental causes of market failure in innovation: knowledge is non-rival (one firm's use does not diminish another's) and partially non-excludable, so inventors cannot capture the full social value of their inventions. Arrow highlighted three key violations of the standard welfare assumptions in the context of R&D: (i) non-convexities in production (the first copy of an idea is costly, but additional copies are cheap, violating convexity of technology), (ii) externalities (innovations create value for others that the inventor cannot fully appropriate, especially as today's ideas become inputs into tomorrow's innovations), and (iii) uncertainty (research outcomes are risky, and with incomplete markets for risk, the outcome is not Pareto efficient)[1]. These insights underpin the "market failure" rationale for policies like patents and R&D subsidies.

While Arrow's analysis centered on incentive problems (and risk), it implicitly involved an *information problem* as well: if knowledge created by one agent benefits others, it means valuable information (the new knowledge) is not fully transmitted or accounted for in market prices. Our work expands on this by explicitly modeling the distribution of information about innovations.

Mechanism design and distributed information. A parallel stream of literature in the mid-20th century studied the allocation of resources under decentralized information. Classic works by Hayek (1945) emphasized that in an economy, knowledge of productive opportunities is dispersed among individuals, and the price system serves as a communication mechanism to aggregate this information. However, when goods have public good characteristics or externalities, prices alone may not convey sufficient information for efficient outcomes. Hurwicz (1960) and Hurwicz (1972) formally founded the field of mechanism design by asking how, in general, one can achieve optimal resource allocation when information (such as preferences or technologies) is privately held by agents. Hurwicz introduced the notion of incentive-compatible mechanisms and highlighted that for certain environments (like public goods), no decentralized mechanism using voluntary communication can achieve the first-best outcome. These insights are directly relevant to innovation: the knowledge of how to innovate or which projects are promising may be privately held, and innovation itself is a kind of public good. Indeed, our model can be viewed as a public goods problem (knowledge creation) complicated by distributed private information about that good.

The failure of the First Welfare Theorem in settings with externalities or public goods is well-understood in mechanism design. The Second Welfare Theorem also fails when the production set is non-convex (as with increasing returns in knowledge) or when information is incomplete. In such cases, one cannot simply assign property rights or redistribute wealth to achieve efficiency — a suitable mechanism must be designed. Our analysis will formalize

[1]See Arrow (1962) for a detailed discussion. Arrow noted that if the price of knowledge were set equal to its (near zero) marginal cost of use, no one would have an incentive to invest in creating it, whereas if knowledge is priced above marginal cost, it will not be disseminated efficiently. He also noted that knowledge spillovers imply that the private benefit of an invention is less than the social benefit, and that uninsurable risk or unobservable effort in R&D creates additional divergence between private and social optima.

this in the context of innovation, echoing the themes of Hurwicz: the structure of information distribution fundamentally limits what market equilibria (or any mechanism) can achieve.

Cumulative innovation and knowledge externalities. Subsequent literature in the economics of innovation has studied how new innovations build on existing knowledge. Scotchmer (1991) used the phrase "standing on the shoulders of giants" to describe how each innovator benefits from prior discoveries. She and others (e.g., Green and Scotchmer, 1995) explored the problem of sequential innovation, where an initial invention enables improvements or follow-up inventions. A core insight is that without coordination, the initial inventor may invest too little, since they may not reap rewards from follow-on improvements. This is a problem of both incentives and information: later innovators might have knowledge of how to improve a basic invention, but if the initial step is not taken, those improvements never occur.

Empirical and theoretical studies have documented various distortions arising in sequential innovation environments. For instance, **?** argue that when innovation is highly sequential and complementary (each invention builds essentially on its predecessors), strong patent protection can paradoxically *reduce* the rate of innovation. The reason is that patents can impede the flow of knowledge: if inventors rely heavily on each others' discoveries, allowing free imitation (knowledge sharing) may spur faster cumulative innovation, whereas strict patents create hold-up problems where each inventor might block others or demand fees, slowing down progress. This underscores that the distribution of knowledge across agents (and their ability to use each others' innovations) is central to innovation performance.

Other important contributions have studied how the market or institutions can coordinate knowledge. **?** proposed that patents serve as a "prospect" that allows an initial inventor to coordinate subsequent R&D and improvements (by licensing or by acting as a gatekeeper), which could mitigate duplicative research and ensure that knowledge is used efficiently. However, if patent rights are fragmented (as in patent thickets), coordination can break down into an "anti-commons" where too many overlapping rights prevent efficient combination of knowledge (**?**). Our work formalizes a related point: if pieces of essential knowledge are held by different agents, a decentralized system may not bring them together optimally, absent an overarching mechanism.

Innovation incentives: patents, prizes, and beyond. There is a large literature comparing various innovation policy instruments. Wright (1983) provided an early formal analysis of patents vs. prizes, highlighting that patents (a market mechanism) and prizes (a government reward) have different advantages: patents let the market determine the value of an innovation but create monopoly distortion, whereas prizes (if the government can correctly estimate the innovation's value) avoid monopoly pricing but require public funds and knowledge of the value. Shavell and Ypma (2001) extended this comparison, showing that when a government's information about an innovation's value is good, prizes can achieve higher welfare by eliminating deadweight loss, but if the government is poorly informed, patents might be preferable as they rely on decentralized market valuation. Both of these works largely assume that the main issue is providing incentives to a single innovator or a sequence of innovators, rather than the aggregation of knowledge from multiple sources.

In practice, various hybrid mechanisms have been proposed to harness both market information and public-good advantages. **?** suggested patent buyouts (the government buys a patent at a price determined perhaps by an auction, then puts the innovation in the public domain), attempting to combine market valuation with elimination of monopoly pricing. Advance market commitments (AMCs), as discussed by Kremer (2000), promise a guaranteed market (at a pre-specified price and quantity) for certain innovations (notably vaccines for diseases prevalent in low-income countries). An AMC is effectively a targeted prize paid per unit sold, which both assures innovators of revenue and ensures that the innovation, once made, is distributed widely at low cost. We will include AMCs in our analysis as an example of a mechanism that addresses a specific type of market failure (lack of market demand due to poverty, which is another friction) while also potentially aggregating information about the feasibility of a product (since firms will respond only if they privately believe they can meet the target).

Finally, economists have also examined the role of open science, collaboration, and information-sharing institutions. For example, open-source software development and certain research consortia rely on decentralized contributions without strong intellectual property, instead using reputation or relational incentives. While our formal analysis centers on patents, prizes, and AMCs, our findings about information aggregation shed light on why open collaborative approaches can sometimes outperform market-based ones: when information is widely dispersed, mechanisms that encourage sharing and pooling of knowledge can be more effective.

In summary, prior literature provides pieces of the puzzle: the public good nature of knowledge (Arrow), the need for mechanism design under dispersed information (Hurwicz), the challenges of cumulative innovation (Scotchmer, Bessen-Maskin), and the trade-offs of different incentive schemes (patents vs prizes). This paper synthesizes and builds upon these insights by explicitly modeling distributed information in the innovation process and examining the interplay between information aggregation and incentives in determining innovation outcomes.

3 Modeling Innovation with Distributed Information

In this section, we develop a formal model of innovation where knowledge is cumulative and information about existing knowledge is distributed among agents. The model is kept intentionally simple to focus on the key friction: no single agent has all the information (or all the pieces of knowledge) necessary for innovation, and the market does not automatically price or aggregate those pieces. We first describe the environment, then define the innovation production function and information structure, and finally discuss equilibrium concepts.

3.1 Environment and Agents

Consider an economy with N agents (indexed by $i = 1, 2, \ldots, N$) who can engage in research and innovation. There is a continuum of consumption goods and a numeraire good (money). Each agent is risk-neutral for simplicity and has access to a research technology. We will initially describe a two-period model to illustrate sequential innovation, and then discuss how the insights extend to multiple periods or continuous time.

Time periods: There are two periods (Period 1 and Period 2). In Period 1, agents have the opportunity to undertake a first-generation innovation project. If a project in period 1 is successful, it produces a new piece of knowledge (or an intermediate innovation). In Period 2, agents can undertake a second-generation innovation project that builds on the results of Period 1. After Period 2, payoffs are realized and the game ends. (One could consider a longer horizon with more sequential innovations; the two-period case captures the essence that today's innovation becomes input for tomorrow's.)

Innovation projects: In Period 1, there is a potential project (call it Project A) that requires effort or investment. We assume for now that at most one successful Project A can be realized (agents might race or coordinate, as described below). If Project A succeeds, it creates a new idea or prototype which has two kinds of value:

- It yields an immediate private value V_A if exploited commercially (e.g., through a product or process improvement that can be used/sold in the market).
- It serves as a necessary input for a follow-on project in Period 2 (call this Project B). Without the knowledge from A, Project B cannot be undertaken (or has a much lower probability of success).

In Period 2, Project B (the follow-on innovation) can be attempted, but it can only succeed if Project A was successful and its knowledge is accessible. If successful, Project B yields a private value V_B in commercial terms. We also assume both projects, if successful, create broader social value by improving consumer welfare or providing knowledge spillovers beyond the immediate V_A, V_B (we will account for that in the planner's problem, but individual firms only care about V_A, V_B which they can appropriate through sales or use).

For simplicity, assume each project requires an investment cost or effort c (assumed equal for both projects for now) and the probability of success is p if the investment is made (we could allow different probabilities p_A, p_B and costs c_A, c_B, but it doesn't change qualitatively the analysis). We assume risk-neutrality, so expected payoffs suffice.

3.2 Distributed Information Structure

The critical feature of the model is that knowledge and information are initially dispersed:

- Different agents may have different *pieces of knowledge* or skills needed to implement Project A. For instance, one agent might know one technique, another agent knows another; only by combining these can the project succeed. Alternatively, one agent might simply have an idea that Project A is promising (a private signal of a high success probability p), whereas others are pessimistic.
- Likewise, the ability to carry out Project B, or knowledge of its potential value V_B, might reside with a different agent than the one who can do A. For instance, Agent 1 might be capable of the foundational research in Project A, while Agent 2 knows how to take that result and turn it into an applied innovation B.

We model this in a stylized way: suppose Agent 1 is the only one who can attempt Project A (or has a clear lead in doing so), and Agent 2 is the only one who can effectively carry out Project B. This captures extreme specialization of knowledge. More generally, we could allow any agent to attempt either project, but each has private information about the likelihood of success or cost of each project for themselves. What is important is that no single agent can both do A and B with equal efficiency: e.g., Agent 1 might have an efficiency advantage in A, Agent 2 in B.

To formalize, let θ_1 be a parameter known only to Agent 1 that affects Project A's payoff or probability (for example, θ_1 could make the success probability $p(\theta_1)$ or the cost $c(\theta_1)$). Similarly, let θ_2 be private information of Agent 2 relevant for Project B (assuming Project A succeeds). We can think of θ_1 as summarizing Agent 1's knowledge about how promising A is, and θ_2 as Agent 2's knowledge about the potential of B or how to implement B. These θ_i are independent pieces of information; no one except Agent i observes θ_i initially. There is no public revelation of θ_1 unless Agent 1 undertakes A and perhaps shares the results. If A is successful, some information becomes public: specifically, the knowledge from A itself (the scientific or technical insight) might become available, at least to Agent 1 and possibly to others. We assume that if A is successful, Agent 1 can, if desired, communicate the knowledge to Agent 2 (e.g., by publishing or via a transaction). However, absent any mechanism, Agent 1 might not want to share it freely, and Agent 2 cannot proceed with B without it.

This environment encapsulates an *information friction*: the knowledge enabling B (output of A) is not automatically available to the agent who needs it. It must be transmitted, either via market transactions or some institution. Also, ex ante, the promise of B's value V_B is privately known to Agent 2; Agent 1 might not fully know how valuable A's knowledge would be to B.

3.3 Social Optimum

Before analyzing decentralized outcomes, consider the first-best social optimum. A social planner who knows all information (θ_1, θ_2, and the functional relationships) and can dictate actions would solve:

$$\max \mathbb{E}[\text{Total Social Surplus}] = \mathbb{E}[W],$$

where W includes the net benefits of any innovations, including externalities or consumer surplus.

In our two-period example, the planner would choose whether to invest in Project A in period 1 and Project B in period 2. If θ_1 and θ_2 suggest that the expected $V_A + V_B$ (plus any external social benefits) exceeds the costs, the planner would have both projects undertaken. The social value of Project A is not just V_A but V_A+ (the incremental increase in probability or value of Project B being successful, V_B). In effect, because A enables B, the combined value might be superadditive.

For instance, suppose success probabilities and values are such that: - If A is done (success probability p) and succeeds, then B can be done in period 2 with success probability p and yield V_B. - If A is not done, B cannot be done at all (success probability 0). Then the social planner's expected value from undertaking both projects (with appropriate timing) is

$p \cdot V_A + p^2 \cdot V_B$ (since B yields V_B only if A succeeded). The total cost is c for each project attempted. The planner would undertake both if

$$pV_A + p^2 V_B \geq c_A + c_B.$$

We will assume parameters such that indeed the socially optimal strategy is to attempt A in period 1 and, conditional on success, attempt B in period 2. This yields the highest expected social payoff.

We note that the social planner can also ensure that knowledge flows freely: if A succeeds, the planner will make sure that the knowledge is available to whoever does B (since the planner's objective is total surplus, not individual profits).

3.4 Decentralized Outcomes without Coordination Mechanisms

Now consider how a laissez-faire market would function in this environment, *absent any special mechanism like patents or prizes*. We assume that in the market: - If an agent undertakes a project and succeeds, they can earn the private payoff V associated with that project (by selling the product or using it internally). However, they cannot automatically charge others for the spillover of knowledge, because knowledge is not excludable in the baseline scenario (no IP rights yet). - There is no market for "ideas" per se. That is, Agent 1 cannot sell the idea or partial results of A to Agent 2 because any attempt to do so would require revealing the idea (which, by Arrow's information paradox, would give Agent 2 the information without payment) in the absence of IP protection. - Each agent acts individually to maximize their expected profit, taking others' actions as given (Nash equilibrium concept).

In our two-agent, two-period story: - In period 1, Agent 1 will decide whether to invest in Project A. Agent 2 is not directly involved in period 1 (since only 1 can do A by assumption). - In period 2, if Project A succeeded and if the knowledge from A is public or accessible, Agent 2 would decide whether to invest in B.

Crucially, consider Agent 1's decision in period 1. If Agent 1 invests in A and succeeds, they get V_A (private return from A). They do not internalize V_B because B will be done by Agent 2 and Agent 1 cannot charge Agent 2 for using A's knowledge (without a mechanism). Agent 1 incurs cost c for A. Therefore, Agent 1's expected payoff from doing A is:

$$\Pi_1 = p \cdot V_A - c,$$

since with probability p it succeeds and yields V_A. (We assume if A fails, nothing happens and the cost is still incurred.)

Agent 1 will undertake A if $\Pi_1 \geq 0$, i.e., $pV_A \geq c$. If this condition fails (i.e., the private benefit of A is less than its cost), Agent 1 will not do A, even if V_B is huge, because V_B accrues only to Agent 2.

Agent 2's decision in period 2: If A was not done, B cannot be done. If A was done and succeeded, the knowledge is presumably available (we might assume knowledge eventually leaks or becomes public by period 2, or at least Agent 2 can acquire it at negligible cost once it's created if there's no IP). So Agent 2, in period 2, if A succeeded, will do B if $pV_B \geq c$ (assuming Agent 2's cost and probability are similar form). If that holds, Agent 2 goes ahead, gets $\Pi_2 = pV_B - c$ (nonnegative by assumption if they proceed).

From a social perspective, the condition to get both A and B was $pV_A + p^2V_B \geq c + c$ (because B yields only if A succeeded). We can see a potential gap: it is possible that

$$pV_A < c,$$

so that Agent 1 won't do A, but

$$pV_A + p^2V_B > c + c,$$

so that socially it would be worthwhile to do both. In fact, even if $V_A = 0$ (the first innovation has no immediate private value but is purely an enabling step), it might be worth doing if p^2V_B (the eventual value of the follow-on) exceeds the total costs. But a private actor with $V_A = 0$ would never undertake A because they get no reward.

This simple analysis already indicates a **failure of the First Welfare Theorem**: the competitive (market) outcome where each agent pursues their own profit can be inefficient due to a missing market for the contribution of Project A to Project B. The knowledge that A produces has social value as an input to B, but it is not priced in the market, so Agent 1 cannot profit from that value and therefore underinvests.

We formalize these insights in the next section, and then introduce mechanisms to try to fix the inefficiency.

4 Failure of the Welfare Theorems in the Presence of Information Frictions

In a standard Arrow-Debreu economy with complete markets and no externalities, any competitive equilibrium is Pareto efficient (First Welfare Theorem), and any Pareto efficient allocation can be supported by some competitive equilibrium given appropriate transfers (Second Welfare Theorem). Our innovation economy violates the conditions of these theorems in multiple ways. We highlight two major issues: non-excludable inputs (knowledge externalities) and distributed private information. We provide formal propositions to demonstrate the breakdown of the welfare theorems.

4.1 First Welfare Theorem Failure

Proposition 4.1. *In the innovation economy with distributed knowledge, any decentralized market equilibrium (with no intervention) is generally* not *Pareto efficient. In particular, there exist environments (parameter values and information distributions) such that the competitive equilibrium yields no innovation, even though a Pareto improvement exists where the innovation is undertaken.*

Proof. The example in Section 3.4 effectively serves as a proof by construction. Take the two-period, two-agent model with parameters such that:

- $pV_A < c$ (Agent 1 finds Project A unprofitable privately),
- $pV_A + p^2V_B > 2c$ (the social planner would undertake both projects).

One concrete set of numbers: let $p = 1$ (for simplicity, assume certainty of success), cost $c = 100$ for each project. Let $V_A = 50$ and $V_B = 300$. Then: - Agent 1's private benefit from A is 50, cost is 100; they will not invest. - Social benefit if both A and B are done: $V_A + V_B = 350$, total cost 200, net +150, so it is socially efficient to do both. Yet, in the laissez-faire equilibrium, Agent 1 does nothing. Thus no innovation occurs at all, which is Pareto dominated by the outcome where both innovate (in that outcome, both agents could be made better off—for instance, via a transfer payment from Agent 2 to Agent 1 or a subsidy, since the total surplus is higher by 150).

The failure here is due to a missing market/externality: Agent 1 cannot capture the V_B that their innovation would enable. Therefore the First Welfare Theorem fails because the assumption of no externalities is violated. The innovative knowledge from Project A is an input into B that is not traded; hence prices do not reflect its value to others. As a result, the competitive outcome (no one does A) is not Pareto optimal.

Formally, one can note that the set of commodities in the market does not include "knowledge inputs for B". If we attempted to include it, we would see it is a public good (non-rival usage for B and beyond) and non-excludable without policy intervention. This kind of missing market leads directly to a violation of Pareto efficiency. □

Discussion: Proposition 4.1 underscores the point that knowledge externalities cause under-investment in innovation. This is a well-known result in economic theory of R&D, but our framework emphasizes the role of information distribution: the inefficiency arises because the information (knowledge from A) is produced but not automatically shared in a way that compensates the producer. If Agent 1 and Agent 2 could sign an enforceable contract before A is done, where Agent 2 agrees to pay Agent 1 for enabling B, they might internalize the externality. However, such a contract is impeded by information problems: Agent 1 might not be able to convince Agent 2 of the potential of A without revealing too much (Arrow's information paradox), and they may not even know each other or trust that B will indeed be successful.

Thus, a deeper interpretation is that the market fails to aggregate the information and capabilities of the two agents. In a world with perfect information and complete contracts, one could imagine Agent 2 hiring or financing Agent 1 to do A, because Agent 2 knows A is needed for B and B is valuable. But with private information (θ_1 and θ_2 unknown to the other), such coordination might not occur. The price system alone does not solve it, because there is no price for the "chance to enable B".

4.2 Second Welfare Theorem Failure

One might wonder if the inefficiency could be solved by appropriate transfers or assignment of property rights before innovation starts. The Second Welfare Theorem suggests that if we could redistribute wealth or assign initial endowments suitably, the market might achieve the desired outcome. However, in our setting, the planner cannot simply assign a property right on the yet-to-be-discovered knowledge or lump-sum transfer the value of V_B to Agent 1, because θ_2 (which determines V_B) is Agent 2's private information. Moreover, due to non-convexity (the fixed cost c to produce a non-rival good), the equilibrium might not exist or might not be unique without interventions.

Proposition 4.2. *In the presence of non-convexities (fixed costs for knowledge creation) and incomplete information, the Second Fundamental Theorem of Welfare Economics fails. There exist Pareto optimal allocations (e.g., the one where both innovations are undertaken) that cannot be decentralized as a competitive equilibrium with any assignment of property rights or lump-sum transfers, unless a mechanism is introduced that directly addresses the knowledge externality.*

Sketch of Proof. We outline two key obstacles to decentralizing the first-best via pure market forces:

1. *Non-convexity:* The production set for knowledge is non-convex due to the fixed cost and zero marginal cost nature of ideas. In our example, the creation of knowledge A has a fixed cost c and then can be used freely by B. In a Walrasian equilibrium, if knowledge were a commodity, its price would have to equal marginal cost (which is effectively 0 once created) to get efficient usage. But at price 0, no firm would supply it. This is a classic case where a competitive equilibrium might not exist or fails to support the optimal allocation. No lump-sum transfer can circumvent the fact that if the price of knowledge is zero, a private firm cannot recoup the cost c from sales. One might attempt to allocate the knowledge good as part of initial endowment (i.e., pretend someone initially "owns" the yet-to-be-discovered idea), but that is nonsensical without the idea actually existing. Thus, the condition of convex production sets required for the Second Welfare Theorem is violated.

2. *Private information and missing markets:* Even if we set aside non-convexity, consider trying to achieve the optimal outcome through some property rights assignment. Suppose the social optimum is that A and B are done. For a competitive equilibrium, imagine we give Agent 1 a property right in the knowledge that A would produce, i.e., a right to charge others (Agent 2) for using it. In effect, this is like a patent on A's knowledge, which we will formally introduce later. But without such a mechanism, initially no one has that right. Alternatively, one might try to merge the agents (imagine one company that could do both A and B) by an initial endowment transfer: but since θ_1 and θ_2 are private, we don't know which agent would be better to put in charge, and any such reassignment is equivalent to designing a mechanism.

   More formally, the second welfare theorem typically requires that the planner can pick a desired Pareto efficient allocation and find prices and transfers such that each agent, with those transfers, would choose their part of that allocation. In our case, to get Agent 1 to do A, we would need to promise them a transfer that covers the difference between pV_A and c (if $pV_A < c$). But that transfer would presumably have to come from Agent 2 (or consumers) who benefit from B. If Agent 2 could commit to paying that transfer, it would have done so voluntarily only if it knew B's success value and was assured the knowledge. However, Agent 2 cannot commit to pay Agent 1 ex post without an enforceable contract, and ex ante any such transfer scheme is not in place. The planner cannot just give Agent 1 extra money and expect them to do A unless that money is tied to doing A (otherwise it's not incentive compatible).

   In essence, implementing the first-best requires a contingent contract: "if you (Agent 1) create A, you will receive X from Agent 2 (or from society)". This is not a lump-sum transfer but a mechanism contingent on action and outcome. Without designing such contingencies (which is outside the scope of pure Walrasian equilibrium), the outcome can't be achieved.

Therefore, the only way to achieve the Pareto optimal innovation outcome is to introduce a mechanism that overcomes these frictions (for example, a patent or prize as we analyze next). Pure price and transfer adjustments cannot do the job. This demonstrates the Second Welfare Theorem does not hold here. □

Proposition 4.2 highlights that the nature of innovation (with its fixed R&D costs and knowledge spillovers) defies the assumptions needed to decentralize optimal outcomes via markets alone. This is aligned with the insights of mechanism design literature: when goods are public or information is asymmetric, one typically needs more complex institutions than just competitive markets. In our context, any feasible approach to efficiency must involve either subsidizing the innovator, assigning intellectual property rights, or otherwise intervening to create a market for knowledge or to guide innovation decisions.

In the next section, we turn to exactly those kinds of interventions. We will introduce patents, prizes, and other mechanisms into the model and examine to what extent they can restore efficiency or at least improve outcomes, and how they compare especially regarding the information problem of knowledge distribution.

5 Comparing Mechanisms for Innovation Incentives and Information Aggregation

We now analyze how different innovation policy mechanisms perform in our model. We focus on patents, prizes, and advance market commitments (AMCs) as three distinct approaches: - Patents create property rights in innovations, allowing innovators to exclude others from using the knowledge (for a period) unless compensation is paid. This can create a market for knowledge (through licensing) and provides incentives via monopoly profits. - Prizes offer a reward (usually from a government or patron) for achieving a specified innovation, after which the innovation is typically made public. This separates the reward from the pricing of the end product. - Advance Market Commitments are promises to purchase a certain amount of a product at a given price, conditional on that product being developed. It's a way to simulate a market demand for something that otherwise has insufficient market (often used for socially valuable goods that are not privately profitable).

We will embed each mechanism into our model and derive equilibrium outcomes. Of particular interest is how each mechanism deals with the distributed information: does it help align the two agents (in our example) to collaborate or coordinate? Does it lead to information being revealed or aggregated that would otherwise remain dispersed?

Throughout this section, we maintain the same basic structure as before: Agent 1 potentially does project A and Agent 2 does project B. We now imagine that a policy mechanism is put in place at time 0 (before any actions) which specifies certain rules or payoffs.

5.1 Patent System

Suppose a patent system is in effect. If Agent 1 succeeds in Project A, they can obtain a patent on the knowledge or product from A. This patent gives Agent 1 the exclusive rights to use that knowledge commercially for a length of time (say it covers at least the period until B would be done). What does this imply in our model?

- If Agent 2 wants to undertake Project B, which inherently uses A's knowledge, Agent 2 would infringe on Agent 1's patent unless Agent 1 permits it (typically via a licensing agreement). - Therefore, if A succeeds and is patented, Agent 2 cannot proceed with B independently; the two agents must come to some agreement. In equilibrium, this will usually involve a transfer payment from Agent 2 to Agent 1 (for access to A's knowledge or a joint venture). - Agent 1, anticipating this, now realizes that if they succeed in A, they can capture not only V_A but also potentially some portion of V_B (by charging Agent 2 for the right to do B, or by doing B themselves if possible). - This means Agent 1's incentive to do A is higher under a patent system than without it.

Let's formalize the bargaining or outcome under patents. We will assume that if A is successful, Agent 1 and Agent 2 bargain over the proceeds of doing B. There are various models of licensing and bargaining; for simplicity, assume Agent 1 has all the bargaining power (this is an extreme but analytically convenient assumption). Then Agent 1 will license the knowledge to Agent 2 for a fee equal to (almost) Agent 2's entire gain from doing B. Agent 2's gain from doing B is V_B (since B yields V_B and presumably post-innovation competition or price equals cost aside from the IP aspect). However, if Agent 1 sets a fee too high, Agent 2 might not do B at all. In a simple model, Agent 1 would charge a license fee L that maximizes their payoff. That would typically be $L = V_B - \epsilon$ (just enough so that Agent 2 still has a tiny incentive to do B, assuming B yields zero profit after paying L).

Thus, Agent 1's expected payoff from doing A under patents becomes:

$$\Pi_1^{patent} = p(V_A + (\text{license fee or profit from B})) - c.$$

If we assume full extraction: license fee $\approx V_B$ (assuming $p = 1$ for B for simplicity here, or use expected value if risky), then

$$\Pi_1^{patent} \approx p(V_A + V_B) - c.$$

Agent 1 will do A if $\Pi_1^{patent} \geq 0$, i.e. if

$$p(V_A + V_B) \geq c.$$

This is a much weaker condition than $pV_A \geq c$ (which was needed without patents). In fact, if V_B is large, this condition might hold even if V_A is tiny or zero. Thus, with patents, Agent 1 is motivated to do A in cases where previously they were not. In our inefficient example earlier, $pV_A < c$ but $p(V_A + V_B) > c$, so now A will indeed occur.

So patents can solve the underinvestment problem in that example. However, patents introduce other distortions: - After A is successful, Agent 1 has monopoly power over that knowledge. In a static sense, if A itself is a product or has uses aside from B, Agent 1 might price it above marginal cost, causing some deadweight loss in period 1's usage of A. (We did not model consumption of A's product explicitly, but V_A presumably came from some monopoly profit or some use; under patents we might say V_A includes monopoly profit on A's product, which could be less than the social value of that product). - More subtly in our model, after A is done, the license negotiation for B could fail or could be inefficient if there is bargaining friction or asymmetric information between Agents 1 and 2. We assumed Agent 1 knows V_B and can extract it. If Agent 1 was uncertain about Agent 2's capability or value for B, they might set a fee that sometimes prevents B from happening (if they overestimate, Agent 2 walks away). This is a potential inefficiency due to information asymmetry in bargaining. - Also, if multiple parties were involved (say multiple complementary patents needed for B), coordination becomes a problem (the anti-commons issue).

For our analysis, let's stick to the broad strokes. We can state:

Proposition 5.1. *A patent system can induce investment in innovations that would not occur under laissez-faire by allowing innovators to internalize the externalities of follow-on innovations. In the two-stage model, if $p(V_A+V_B) \geq c$, the patent system achieves innovation of A (whereas without patents A would require $pV_A \geq c$). However, the patent system may introduce inefficiencies due to monopoly pricing and bargaining problems. In particular, if V_A and V_B are realized outcomes, the patentholder will extract rents that may reduce consumer surplus, and if information about V_B is private to the follow-on innovator, the licensing negotiation may fail to realize B even when socially beneficial.*

Discussion/Proof. The key incentive effect of patents is as derived above: Agent 1's incentive constraint becomes $p(V_A + \alpha V_B) \geq c$, where α is the fraction of follow-on value they expect to capture (with $\alpha = 1$ under full bargaining power for Agent 1). This is strictly larger than the no-patent condition. Thus, patents strictly expand the set of parameter values for which A is undertaken, moving the outcome closer to the first-best. In the extreme of $\alpha = 1$ and risk-neutrality, if the patent duration or breadth is such that V_B can be fully appropriated, then any socially worthwhile A ($pV_A + p^2V_B \geq c + c$) would also be privately worthwhile ($p(V_A + V_B) \geq c$ implies roughly $pV_A + pV_B \geq c$, which for large V_B and decent p will hold if the social condition holds, though not exactly the same condition because of the p^2 term versus p for V_B—this slight difference is because the patentholder might also consider the probability that B succeeds p).

The inefficiencies stem from what happens after A: - If V_A comes from selling a product under monopoly, the social value of that product is higher (because price ¿ marginal cost means some consumers who value it more than cost but less than the monopoly price do not get it). In our framework, we didn't explicitly model consumers, but V_A could be thought of as monopoly profit, whereas the total surplus from A's product might be larger. Patents intentionally sacrifice some static efficiency to reward the inventor. - For project B, under patents, the knowledge is not freely available. The need for licensing can delay B or add transaction costs. If information θ_2 (Agent 2's cost or chance of success) is private, then we have a classic bilateral trade with private information problem (see **?**). Myerson and Satterthwaite's theorem on bilateral trade says that if two parties each have private info about their values, then even with bargaining, you cannot always achieve all gains from trade with a voluntary mechanism. In our context, that implies that sometimes B might not happen even though A succeeded and it would be socially efficient for B to happen, because the parties fail to strike a deal (for example, Agent 1 demands too high a fee given their uncertainty about Agent 2's success probability or cost). Patents don't automatically solve that; they just give a chance to negotiate. - If multiple follow-on innovators or multiple pieces of knowledge from different patent holders are needed for a single follow-on innovation, coordination becomes even more challenging (Heller and Eisenberg's anti-commons problem). That would amplify information frictions, as each might hold out or not know the other's stance.

These points explain why a patent system, while improving the incentive for initial innovation, does not guarantee a first-best outcome in general. Nonetheless, compared to no intervention, patents typically increase the rate of innovation in sequential settings by aligning private and social returns more closely. □

Given these effects, the patent mechanism in our model would succeed in getting Agent 1 to do Project A in scenarios like our example, and then presumably Agent 2 does B after paying a license fee. The overall social outcome with patents is that both innovations occur (which is good), but there is a transfer of surplus from Agent 2 (or consumers of A or B) to Agent 1, and potentially some loss of consumer surplus.

We next turn to prizes, which take a different approach.

5.2 Prizes

Now consider a prize mechanism. Suppose some sponsor (e.g., a government or foundation) offers a prize for the successful completion of Project B (the final innovation). That is, if someone successfully achieves the end goal, they will receive a reward P dollars. We can also consider the possibility of a prize for Project A if needed, but let's start with prize for B only, since B is the ultimate objective.

Under a prize for B: - Agent 2 (or anyone) knows that if they are the first to complete Project B successfully, they get P. - We assume that claiming the prize requires disclosure of the innovation (so that after awarding, the innovation is public domain, which is typical for prize designs). - There is no patenting (or even if patenting is allowed, the prize might stipulate open access; but let's assume it's prize *instead* of patent).

How does this affect the incentives? - Agent 2's incentive to do B becomes: if A is done and available, Agent 2 gets $P + V_B$ by succeeding (they get prize plus any commercial value; though if it's open after prize, maybe V_B as a market value is moot, but maybe assume V_B was consumer benefit or something; let's assume the prize is the main reward). - If A is not done, B can't be done, so irrelevant.

But what about Agent 1 (Project A)? Under a B-prize only, Agent 1 has no direct reward; they might still not do A unless they foresee some benefit. However, consider that Agent 2 cannot do B unless A is done. If Agent 2 is free to collaborate or contract, perhaps Agent 2 would now have incentive to ensure A gets done so that they can go for the prize. Agent 2 might approach Agent 1 and say: "If you do A, I'll pay you something or we can form a team so that we can then do B and I can claim the prize." But if the prize goes only to whoever actually completes B, they'd presumably share it under a team agreement.

In practice, prize competitions often allow team entries, so Agent 1 and Agent 2 could form a team where if they together achieve B, they split the prize according to some arrangement. However, that arrangement is not enforced by the prize giver; it's a private contract. Still, one might expect them to coordinate. The difficulty again is trust and contracting:

Agent 1 might fear that if they do A, Agent 2 could try to cut them out of the prize credit, or vice versa.

If we assume for simplicity that A and B can be achieved by the same entity if they collaborate (like a team acting as a single unit), then effectively a prize for B could also incentivize the combined effort. One extreme case: Agent 2 could undertake A themselves even if they're less efficient, or hire Agent 1. This becomes similar to a scenario of an integrated approach to win the prize.

However, the simplest modeling: consider that the prize designer might also offer an intermediate prize for A or at least recognize that enabling steps matter. In some prize designs, milestones are rewarded.

Let's consider two versions: - Prize only for B: Then likely, to be realistic, Agent 2 would attempt to do both A and B if possible (if Agent 2 can do A at all). If Agent 2 cannot do A at all (only Agent 1 has that capability), then Agent 2 has to persuade Agent 1 to do it. Possibly Agent 2 could promise a side payment from the future prize. But if no contract, Agent 1 won't trust that. So a pure B prize might fail to elicit A if the knowledge for A is separated. - Prize for both A and B: Suppose the sponsor offers prize P_A for achieving A (perhaps verifying some intermediate result) and prize P_B for achieving B. This covers both steps. Then Agent 1 could go for P_A, and Agent 2 for P_B (with the assumption that after A's prize is won, A's knowledge is published as part of prize verification, so Agent 2 can use it openly). This could solve the issue: Agent 1 gets reward for A, so they do it if $P_A \geq c$ (assuming success certain for simplicity). Agent 2 does B if $P_B \geq c$. The prizes would have to be funded, presumably set equal to the social value of each step or so to induce participation.

However, paying for intermediate steps requires the sponsor to know what intermediate is needed and verify its success. In many cases, sponsors only care about final outcomes, not how achieved.

Given the focus, let's state results in general terms:

Proposition 5.2. *A prize system (rewarding innovation outcomes with public funds) can in theory achieve the first-best innovation outcome, by setting appropriate prize amounts that reflect the social value of innovations. In particular, if a prize $P_B = pV_B$ is paid for successful completion of project B (the follow-on innovation), and if knowledge from project A is freely disseminated, then a planner could also offer a prize $P_A = pV_A^*$ for project A or otherwise arrange that the team that achieves A and B collectively receives $P_A + P_B = p(V_A + V_B^*)$, where V_A^*, V_B^* denote the social values (including spillovers). By choosing prizes equal to social marginal values, the planner can induce agents to undertake both projects exactly when it is socially optimal. Furthermore, because the innovation is made public after the prize, there is no monopoly distortion in usage of the knowledge. However, the challenge for prize systems is that the sponsor must know or determine the appropriate prize value, and coordination between agents may still be needed if tasks are separated.*

Discussion. The proposition is a bit informally stated, but it captures known results in the literature such as Shavell and Ypma (2001): if the government knows the social value of an innovation, it can pay a prize equal to that value to induce an agent to produce it, and then make it freely available, achieving the first-best outcome. In our context, the complication is two-stage.

To get Agent 1 to do A, the prize for A would need to cover the gap in private incentives. If only a prize for B is offered, one must rely on the second agent (who aims for the B prize) to engage the first agent's help. If contracting between agents is frictionless, Agent 2 could effectively subcontract Agent 1's work, promising some share of the prize. In reality, without enforceable contracts, one might worry.

For simplicity, assume the sponsor offers both P_A and P_B. Then: - Agent 1 will do A if $pP_A \geq c$ (assuming if they succeed they definitely get P_A). - Once A is done, the knowledge is public (prize competitions often require disclosing the solution). Now Agent 2 (or anyone) can do B. Agent 2 will do B if $pP_B \geq c$. So as long as the prizes are at least as large as the costs (scaled by success probability), both happen. - The social optimality requires not overspending either; ideally P_A and P_B are set equal to the expected marginal social benefit each provides. If set correctly, the outcome is first-best: both projects occur if and only if socially beneficial, and the results are public, maximizing usage.

No monopoly pricing occurs, since after the prize, the innovation is in public domain or at least priced at marginal cost (0 for knowledge).

The downside: the sponsor needs to know the right prize values. In practice, this is hard because it requires information on V_A and V_B which are privately known. So there's an information problem at the level of mechanism design: the sponsor might guess or use some mechanism (like an auction or contest among entrants) to gauge the necessary reward.

Coordination: If only a prize for B is given, theoretically any team can claim it. So Agent 1 and Agent 2 could form a team to claim the prize for B together, splitting the reward internally. That is essentially equivalent to having an implicit prize for A (part of B's prize goes to the person who did A). Prizes historically have led to collaborations (e.g., the Ansari X Prize for spaceflight had teams pooling multiple talents). One risk is that without clear rules, team formation is itself a negotiation problem.

However, unlike patents, prizes do not legally constrain use of knowledge. This means if A is done by someone and not shared, others might independently redo A in order to get to B prize. But rationally, if someone succeeded in A, they'd publish or announce it if they are going for B prize themselves (unless they think they can keep it secret and get B prize solely, but they might need help). In any case, because the prize requires revealing the solution, knowledge tends to disseminate more than under patents (where it might be kept secret if not patented, or disclosed in patent but not freely usable).

Therefore, an ideal prize mechanism can overcome both incentive and usage inefficiencies, but it hinges on the sponsor's information and might require breaking the problem into stages if there are distinct steps. □

In summary, a prize regime can potentially achieve what patents cannot: no deadweight loss in knowledge usage, and no need to grant monopoly. But it requires the difficult task of knowing the value of innovation to set the prize, or otherwise ensuring that competition for the prize drives efficient outcomes.

From an information perspective, prizes are somewhat top-down: they do not automatically aggregate decentralized knowledge unless participants choose to collaborate. In our model, if Agent 1 and Agent 2 don't team up, Agent 2 might try to do A themselves even if they are less efficient, leading to duplication or delays. So one might see multiple agents trying the whole task independently, which can be both good (parallel experimentation) and bad (duplication of effort). In contrast, patents naturally allow specialization: Agent 1 does A, then Agent 2 does B after licensing. Prizes might induce each competitor to vertically integrate to win. Whether that is efficient depends on context (see **?** for analysis of prize contests with sequential innovations, for example).

5.3 Advance Market Commitments (AMCs)

Advance Market Commitments are a mechanism originally proposed to incentivize vaccine development for diseases affecting low-income countries. An AMC typically works as follows: donors commit a fund that will subsidize the purchase of a vaccine (or other product) up to a certain amount (both price per unit and total quantity) if a vaccine meeting certain specifications is developed by some date. For instance, they might guarantee to buy (or subsidize) 200 million doses at $5 each for a vaccine for disease X.

We can interpret an AMC in our model terms. Suppose the final product of Project B is a vaccine that has low market value because the consumers are poor. So V_B (the private profit from selling it in the market) might be low, but the social value (health benefits) is high. An AMC essentially raises V_B artificially by injecting donor money. Instead of a fixed prize after development, it is a commitment to pay per unit, but from the developer's perspective it is a reward for success with an implicit expectation of production.

In our model, how does an AMC affect the game? - If an AMC is available for B, then Agent 2 knows that if they succeed in B, they will be able to sell a certain amount at a good price. This increases the expected V_B (private) for them, potentially from something negligible to something significant. - So it directly incentivizes Project B similarly to a prize, but with the nuance that it ensures the product reaches users (since the payout is tied to units sold or used). - Agent 1's issue with A remains similar: if Agent 2 stands to gain more from B (thanks to AMC), Agent 2 might be willing to contract or partner with Agent 1 to ensure A is done. Or, if patents exist concurrently with AMC, Agent 1 could patent A and then license to Agent 2 who will make money from AMC-backed sales. - If no patents and just AMC, we rely on maybe a single firm tackling both or collaboration.

A difference between prize and AMC is that AMC typically doesn't require the product to be non-excludable after. In fact, the developer might still have a patent but is selling to the AMC at an agreed price. However, often AMC deals involve the developer agreeing to supply at an affordable price beyond the subsidized amount, etc. But for theory, it's essentially a guarantee of a market at a high price for a limited quantity.

One could model AMC as: If B is done, the inventor gets $\tilde{V}_B$ in profit instead of V_B, where $\tilde{V}_B > V_B$ due to the subsidy. It's like a prize that is proportional to usage.

For information aggregation: The AMC might specify desired product characteristics but not how to achieve them. It leaves it to firms to figure out. Similar to prizes, multiple firms might try and they might not share info. However, sometimes AMC are done after some partial info is known or with collaboration with health orgs.

We can articulate:

Proposition 5.3. *An Advance Market Commitment (AMC) increases the private returns to successful innovation by guaranteeing a paying market. In our model, an AMC aimed at the final innovation (B) effectively boosts V_B (the private reward for B) to a higher level $\tilde{V}_B$.*

This can ensure that even socially valuable innovations with low immediate market value are pursued. Compared to a lump-sum prize, an AMC has the advantage of leveraging market mechanisms for distribution (rewarding per unit sales ensures the product is produced and delivered). However, similar coordination issues arise: if different agents control different parts of the innovation chain, the AMC alone does not solve how knowledge from A reaches the developer of B (unless one entity undertakes both or collaboration occurs). In combination with patents or contracts, AMCs can facilitate coordination by providing a larger pie for parties to share.

The proof/argument is straightforward given the above discussion. AMC can be thought of as a variant of prize that pays out in a way tied to output, which might reduce the risk of over- or under-estimating the value (as it pays only if there's demand).

One could also note: if an AMC is structured poorly, it could lead to rewarding a product that isn't the very best possible (if technology moves on, etc., but that's detail).

In our two-agent story, an AMC might cause Agent 2 to be extremely interested in doing B. If patents are present, Agent 2 will likely pay Agent 1 for A (since now the follow-on value is high, easier to strike a deal). If patents are not present, maybe Agent 2 hires Agent 1 or works together. In both cases, the prospect of AMC money helps bring them together, although one might still see a scenario without patent where Agent 1 hesitates unless Agent 2 credibly shares the AMC reward. Possibly Agent 2 could pay some advance to Agent 1 from investors, knowing AMC will pay back later.

We can mention that mechanisms not analyzed in depth but relevant: government R&D grants (which pay for A directly, essentially removing the need for Agent 1 to consider profit; the government just funds it). That is akin to a prize but paid upfront or cost-sharing.

Also "open source" style: no patents, but rely on intrinsic or alternative incentives; not directly in our formal analysis, but conceptually, if many contribute knowledge freely (like in some software contexts), initial knowledge is provided voluntarily, enabling follow-ons, and often supported by alternative reward systems (like reputation or later commercialization of complementary goods).

Given the scope, let's proceed to discussion.

5.4 Formal Propositions Summary and Comparison

To compare mechanisms more directly, we can consider specific criteria: - Does the mechanism achieve the socially optimal innovation investment (incentive efficiency)? - Does it ensure the knowledge is widely used (allocative efficiency)? - How does it handle dispersed information (coordination of multiple innovators)?

We summarize insights from the above analysis in the following comparative propositions:

Proposition 5.4. *In the sequential innovation model:*

(a) ***Patents vs. No intervention:*** *Patents weakly increase the equilibrium amount of innovation by allowing innovators to appropriate follow-on value. Any equilibrium without patents is achievable with patents (through appropriate licensing) and patents induce additional innovation in cases of significant externalities. However, patents introduce monopoly distortion in knowledge usage.*

(b) ***Prizes vs. Patents:*** *If the prize granter has full information about the social value of innovation, an optimally designed prize mechanism can achieve the first-best outcome, which dominates the best achievable patent outcome in terms of total welfare (because it avoids deadweight loss). However, if the prize granter is uncertain about the value, patents may lead to better selection of projects (the market reveals value). Furthermore, prizes may lead to parallel redundant efforts, whereas patents concentrate effort but risk underutilization of knowledge without licensing.*

(c) ***AMCs vs. Prizes:*** *An AMC for a final product can achieve similar outcomes to a prize for that product in terms of incentives, but has the advantage of tying rewards to actual usage (mitigating the risk of overpaying for a useless innovation). In terms of information, both rely on competition among innovators rather than explicit information sharing; however, an AMC might encourage firms to invest in learning about consumer needs and production scale-up earlier.*

(d) ***Distributed knowledge aggregation:*** *Patents encourage ex post* transactions *(like licensing, acquisitions) that can aggregate knowledge held by different parties, because owning IP rights gives the ability to trade knowledge. Prizes encourage ex ante* team formation *or integration, since the reward goes to the first to achieve the result regardless of how many parties collaborate internally. In a situation with highly fragmented knowledge, a patent system might fail if too many patent holders create a thicket (coordination costs high), whereas a prize might fail if no single team can assemble all pieces (coordination failure in team formation). A combination, such as a patent with policies to mitigate anti-commons (e.g., requiring licensing on fair terms) or a prize that explicitly rewards intermediate contributions, may perform better.*

The above proposition is a qualitative summary. We could prove parts of it by referencing earlier results: (a) follows from Proposition 5.1; (b) follows from comparing social welfare under patents vs prize (standard results, e.g., Shavell and Ypma, 2001); (c) is more intuitive, referencing Kremer (2000) arguments; (d) is reasoning from our model structure.

Finally, we note that in mechanism design terms, none of these simple mechanisms fully solve the problem of distributed private information. They each have strengths and weaknesses in managing information: - A patent system delegates decisions to the market: each agent uses their private info to decide what to research. This harnesses local knowledge (agents know their own θ_i) but because of externalities, not all info is used. Patents partly internalize externalities, but if θ_1 and θ_2 are separate, an inefficiency remains unless negotiation happens. - A prize system centralizes the target but not the process: the principal doesn't know θ_i, they just set a goal and reward. Agents then self-select. If θ_1 is favorable, maybe Agent 1 tries for the prize, otherwise not. There's a risk the principal sets wrong targets or amounts due to lack of info. - In theory, one could design a more complex mechanism that asks agents to report signals θ_i and then assigns tasks or rewards accordingly (like an auction for research contracts). That goes into the realm of mechanism design for R&D, which is beyond our current scope but conceptually possible.

The key takeaway is that information frictions mean no simple decentralized mechanism is guaranteed efficient: some knowledge will either be underused or some rent will be necessary to coax it out.

6 Discussion

Our formal results highlight that innovation policy cannot be solely about incentivizing effort; it must also consider the allocation of knowledge and information. In practice, we observe a variety of institutions that complement patents and prizes to handle information issues. For example:

- **Research Joint Ventures (RJVs):** Sometimes firms form alliances to share know-how and jointly develop technologies, partially overcoming knowledge fragmentation. This can be seen as a private arrangement to aggregate information when patents alone would lead to costly licensing or duplication.

- **Patent Pools and Standards:** In industries with many complementary patents (like telecommunications), patent pools allow multiple patent holders to license their technologies as a bundle, reducing transaction costs. Standard-setting organizations also help coordinate which technologies are used, after which patent licensing can be streamlined. These mitigate the anti-commons effect.

- **Open Science and Open Source:** In academia and open-source software, knowledge is openly shared rather than protected. This model relies on alternative incentives (reputation, future career rewards, or indirect commercial benefits) but excels in rapidly disseminating information. Our theory suggests that when the distribution of knowledge is a bigger hurdle than motivation, open approaches can yield faster cumulative innovation.

- **Hybrid Incentives:** In practice, we often see hybrids: for instance, a company might get a patent (securing incentive) but the research that led to the invention might have been funded by a government grant (addressing the early-stage info externality). Or a prize might be offered for a general goal while allowing winners to patent their solutions (as sometimes in defense contracts). Combining mechanisms can sometimes capture the benefits of each.

One interesting implication of our model is how it relates to the current digital age: information about existing knowledge is more accessible than ever (via publications, the internet), potentially reducing some information frictions. However, the sheer volume of knowledge means individual innovators still specialize, so the problem of integrating disparate pieces remains. Our theory would predict that fields where knowledge is very fragmented and cumulative (like complex technologies with many components) might benefit from more collaborative approaches or stronger coordination mechanisms, whereas fields where single breakthroughs can be made by isolated inventors might be fine with patents or small prizes.

Another aspect is dynamic efficiency vs static efficiency: We showed patents cause static inefficiency (monopoly pricing), whereas prizes and open dissemination avoid that. But dynamic efficiency (the rate of innovation) might suffer if we remove patents without an adequate alternative incentive. There's a trade-off and our model quantifies some of that trade: e.g., Bessen and Maskin's scenario where too strong patents can actually slow cumulative innovation, implies that in some cases a regime with freer use of knowledge (hence some static efficiency gain and faster follow-on) yields more innovation in the long run.

Limitations of the model: Our formal model was simplified to two periods and two agents for clarity. Real innovation systems involve many agents, multiple overlapping generations of innovation, competition (multiple entities may attempt the same innovation), and uncertainty in more complex forms. Additionally, we treated information as exogenous private signals. In reality, firms can invest in learning or research that gradually uncovers information. Extending the model to more periods could incorporate R&D races or learning over time. Nonetheless, the qualitative insights about information distribution and coordination should carry over.

Another limitation is that we did not explicitly model consumers and welfare from consumption of innovations in detail. A more complete general equilibrium model could integrate consumer surplus and thereby explicitly account for monopoly distortions. We relied on known results to discuss those qualitatively.

Finally, while we compared patents, prizes, and AMCs, there are other mechanisms like licensing contests, subsidies, or even regulatory tools (e.g., requiring data sharing) that could be analyzed. Future work could also explore how the optimal mechanism might be designed when the planner has some information about the distribution of θ_i but not exact values—a mechanism design problem for innovation incentives under asymmetric information.

7 Conclusion

This paper has developed a formal economic theory highlighting the role of information frictions in the innovation process. Our model shows that when knowledge is a cumulative input to further innovation and is distributed across different actors, market outcomes will generally be inefficient: crucial knowledge inputs are not paid for, and decentralized decisions fail to coordinate on socially beneficial innovation paths. We demonstrated formally why the fundamental welfare theorems do not hold in this setting, due to non-convexities and missing markets for information.

We then analyzed how various mechanisms—patents, prizes, and advance market commitments—perform in this environment. We provided propositions and proofs showing that each mechanism has distinct impacts on both incentives and information aggregation. Patents help align private incentives with social value by creating a market for knowledge, but they risk throwing sand in the gears of knowledge dissemination. Prizes can, in principle, achieve first-best outcomes by using public funds to reward innovations and then making them public, but they require the sponsor to know the innovation's value and to consider the coordination of contributors. Advance Market Commitments offer an interesting hybrid, leveraging market-like rewards for successful innovations while ensuring access.

A key insight from our exploration is that *innovation policy is as much about managing information as it is about providing incentives.* A well-designed innovation system might need to combine elements: for example, patents to allow decentralized discovery of valuable opportunities, but complemented by policies that encourage sharing of knowledge (such as patent pools or mandates for publishing research findings after some time). Likewise, prize designs might consider awarding intermediate milestones to make sure fragmented knowledge holders have incentive to contribute.

Our theoretical framework abstracted from some complexities, but the hope is that it captures the essence of why innovation often requires collective efforts and why purely relying on market forces can lead to suboptimal outcomes. Understanding the information architecture of innovation can help policymakers strike better balances: encouraging innovation while also enabling innovators to "stand on each other's shoulders" rather than reinvent the wheel or guard their secrets too closely.

References

References

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by Richard R. Nelson, 609–625. Princeton, NJ: Princeton University Press.

Green, John R., and Susan Scotchmer. 1995. "A Theory of Cumulative Innovation." *RAND Journal of Economics* 26(1): 3–23.

Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35(4): 519–530.

Hurwicz, Leonid. 1960. "Optimality and Informational Efficiency in Resource Allocation Processes." In *Proceedings of the National Academy of Sciences*, 1586–1591.

Hurwicz, Leonid. 1972. "On Informationally Decentralized Systems." In *Decision and Organization*, edited by C. B. McGuire and R. Radner, 297–336. Chicago: Markham.

Kremer, Michael. 2000. "The ODA Conjecture and the Commercialization of Publicly Funded Innovation." *American Economic Review* 90(4): 1057–1070.

Nelson, Richard R. 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67(3): 297–306.

Scotchmer, Susan. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives* 5(1): 29–41.

Shavell, Steven, and Jef Ypma. 2001. "Contests for Innovative Ideas." *Econometrica* 69(2): 423–449.

Wright, Robert E. 1983. "Patent Policy and Research Strategies." *RAND Journal of Economics* 14(1): 21–37.