If you had only one shot: Scale and herding in innovation experiments

January 13, 2026

Abstract

Solving complex technological problems often requires testing a diverse set of approaches. While conventional wisdom holds that more independent experimenters lead to greater diversity of approaches and higher chances of success, we argue this is not always the case. We develop a simple model demonstrating that, for a given number of experiments, markets dominated by a few multi-experimenters explore more diverse approaches than those with many single-shot experimenters. Single-shot experimenters tend to converge on the most promising approach, while multi-experimenters are more likely to diversify to avoid the correlation inherent in pursuing multiple experiments within the same approach. We test these predictions using data from pharmaceutical R&D. A unit increase in the average experimenter scale is associated with a 0.76 standard deviation increase in the diversity of targets deployed in a therapeutic class–year. Importantly, an increase of one standard deviation in target-diversity is associated with a 26.8 percentage point increase in the probability that a market has at least one launch, but a 3.9 percentage point decrease in the share of projects that succeed in pre-clinical development. Our findings point to the optimal allocation of experiments across firms in a market to maximize market-level success, and the importance of distinguishing individual from market outcomes.

1 Introduction

From healthcare to climate change, solving society's most pressing challenges requires innovation and extensive experimentation. For instance, the invention of the semiconductor transistor arose from systematic, extensive experimentation with materials and design combinations across multiple labs, culminating in the 1947 breakthrough by Bardeen, Brattain, and Shockley. Similarly, Thomas Edison's quest for the ideal filament for the light bulb involved testing thousands of combinations of materials and configurations.

These experiments are not conducted haphazardly; they are guided by specific approaches, or structured hypotheses about how to solve a problem, often designed to explore specific pathways, mechanisms, or combinations of variables in solving a problem (Camuffo et al., 2020; Sorenson, 2024). It is therefore not just the number of experiments that matters, but how many approaches are taken to address the problem. Distinguishing between an approach and an experiment helps disentangle two sources of uncertainty: whether the aproach itself is viable, and whether a given implementation of that approach succeeds.

Breakthroughs often emerge when a range of different approaches are explored For example, effective cholesterol treatment emerged from years of experimentation across multiple approaches, including relative dead ends such as clofibrates and bile acid sequestrants. Success ultimately came with the discovery of statins, which inhibit cholesterol synthesis in the body, based on Akira Endo's groundbreaking work on fermented natural products. During World War II, the US funded multiple rivalrous approaches for the development of radar and the proximity fuse (Gross and Sampat, 2023).

Yet, when innovation is left to market actors, they may converge on similar approaches, even when the problem space allows for diversity. Conventional wisdom holds that many small-scale experimenters will generate more approach diversity than a few large ones running multiple experiments, since independent solvers are assumed to bring independent perspectives (Cohen and Klepper, 1992; Thomke et al., 1998). However, this logic assumes that independence is sufficient to ensure diversity. But even if independent actors choose different

experiments, they may still herd by choosing the same approach.

We develop a simple model to examine how firms decide which approaches to pursue, distinguishing between multi-experiment firms and one-shot experimenters. In our setup, experiments can fail for two reasons: the approach itself may be a dead end, or the implementation may fail even if the approach is viable. Multi-experiment firms reduce the correlation between outcomes by spreading their efforts across different approaches, each with independent viability, thereby increasing the likelihood of at least one success. One-shot experimenters, by contrast, tend to concentrate on the most promising approach, choosing independently but similarly, which reduces diversity at the market level.

Our model departs from traditional herding explanations by showing that even independent decision-making can lead to herding, not because of information cascades (e.g., Bikhchandani et al., 1992), data constraints, or behavioral biases such as homophily, but due to rational responses to incentives. The key intuition is that when only one success matters and payoffs are interdependent, multi-experiment firms hedge against correlated failures by spreading their experiments across multiple approaches, even at the cost of trying less promising approaches. Even with interdependent private signals and no learning, single-experiment firms, on the other hand, rationally herd because the business-stealing motive makes it most profitable to adopt the most promising approach. This results in a tradeoff. Markets dominated by large, multi-experiment firms exhibit greater approach diversity and a higher chance of at least one success, but a lower average success rate per experiment. We assume that the total return to successful firms is fixed and does not increase with the number of winners, capturing settings where society primarily benefits from the first breakthrough.[1]

The model shows how the business-stealing motive can lead small-scale experimenters to pursue common strategies, without the need for learning or updating beliefs. These firms—who are maximising the probability that their one shot succeeds—choose the most promising

[1]For example, most modern machine learning models rely on the transformer architecture, lithium-ion batteries dominate in EVs and electronics, and CMOS has dominated integrated circuits. These are cases where one solution has temporarily captured the field.

approach, ignoring the reduction in rival profits. This incentive pushes single-experimenter firms towards the same approach rather than expanding the diversity of approaches tested.

To test our model's implications, we use a comprehensive dataset on pharmaceutical drug development from 1996 to 2023, covering 49,866 projects led by 3,845 firms across 241 therapeutic classes. The data track drug-target combinations, where a therapeutic class-target pair represents an approach, and a firm's preclinical trial of that pair constitutes an implementation attempt. Our empirical analysis supports our model's predictions. A one-unit increase in the average experimenter scale is associated with a 0.76 standard deviation increase in the diversity of targets deployed in a therapeutic class-year. Importantly, markets with more target diversity are more likely to achieve at least one successful outcome: a one standard deviation increase in target diversity is associated with a 26.8 percentage point increase in the probability that a pre-clinical experiment progresses to Phase 1 clinical trials. However, this benefit comes at the cost of a 3.9 percentage point decrease in the share of projects that succeed in pre-clinical development. These results show that the experimenter scale can foster approach diversity and increase the probability of success in solving complex problems. The results are also robust to controls for the total number of targets, firms, and projects in a therapeutic class, as well as time-varying broad therapeutic area (ATC-1) fixed effects to account for unobserved technological and demand shocks (Branstetter et al., 2022).

The model's core logic is that experimenter scale affects outcomes through its impact on diversity. Our empirical results show that while diversity itself is not systematically related to average success, the component of diversity predicted by experimenter scale *is* negatively related to it. This pattern holds when we control for the rate of new target discoveries—an alternative driver of diversity and success. In fact, when doing so, the negative relationship between diversity and average success becomes stronger, and the positive association with at-least-one success persists.

Other factors could account for some of the patterns we observe. Markets may differ in size, technological opportunity, or firm capabilities, all of which could affect experimenter

scale and the likelihood of success. However, none of these can explain the specific pattern we document: experimenter scale is positively associated with approach diversity, which is positively associated with the probability of at least one successful solution, but negatively associated with average success per experiment. In particular, our results are inconsistent with theories that attribute herding to data availability (the streetlight effect), behavioral factors, or heterogeneity across markets in differences in the superiority of the preferred approaches over others. Notably, only our mechanism predicts that experimenter scale is associated with greater diversity, and that this diversity is associated with the probability of at least one success while lowering the average success rate per experiment.

Empirically, we go beyond the model and explore another source of diversity, namely the discovery of new approaches. While we find that single-experiment firms do indeed introduce a significant proportion of new approaches, this does not diminish the relationship between experimenter scale and diversity. Furthermore, contrary to the assumption that one-shot experiments are exclusive to startups, both large public firms and startups actively engage in single experiments. Nevertheless, small-scale experimenters play a crucial role in introducing new approaches, with firms conducting only one experiment being more likely to explore novel targets than multi-experimenters. These complementary findings suggest that while markets with greater average experimenter scale foster diversity by leveraging the broader experimentation efforts of large-scale firms, they also depend on exploration by small-scale experimenters to discover new targets.

This paper makes three key contributions. The literature on experimentation has primarily focused on the effectiveness of experimentation for firm-level performance (Koning et al., 2022). We shift the focus to market-level outcomes. We show that the scale of experimentation—how many shots each firm takes—shapes market-level diversity and success. This shifts focus from the effectiveness of individual experiments to the collective architecture of experimentation. We thus link experimentation to technology policy, particularly regarding the optimal societal organization of collective experimentation (Nelson,

1961).

Second, we offer a conceptual contribution by distinguishing between two components of an experiment: the approach and the implementation attempt. Implicitly, much of the literature on experimentation, including studies by Koning et al. (2022) and Gans et al. (2019), do not distinguish between the implementation and approach, instead bundling them into a single concept. We argue that these are distinct (Bryan and Lemus, 2017; Dasgupta and Maskin, 1987). An approach refers to a hypothesis or theory about the cause of a problem, suggesting potential ways to solve it (Camuffo et al., 2020; Sorenson, 2024). In contrast, an implementation is a specific attempt to solve the problem based on the hypothesized cause. For success, the approach must be viable *and* the implementation must be effective. As we demonstrate, individual experimenters do not adequately account for the shared risk of failure across all implementation attempts using the same approach.

Third, our empirical results emphasize the distinct ways in which large- and small-scale experimenters affect diversity. Small-scale experimenters contribute by introducing novel approaches, thereby expanding the portfolio of available approaches. Large-scale experimenters, on the other hand, add diversity by more thoroughly investigating known but underexplored approaches. This finding ties to the literature on technological trajectories, suggesting that markets with small-scale experimenters may underexplore some trajectories—herding towards the ones that are most likely to succeed (Ciarli et al., 2021; Nelson et al., 2023; Tan, 2023).

2 Related Literature

In this section, we briefly review the literature relevant to our paper, connecting ideas from three domains: experimentation and its market-level implications, firm-level incentives to innovate, and the relationship between firm size and the diversity of technological approaches. These literatures motivate our model in Section 4, which explores how firm-level experimen-

tation strategies shape market-level outcomes.

The first literature we draw on is the recent work on experimentation, which has thus far focused on the design and effectiveness of experiments, using a decision-theoretic approach. For example, Koning et al. (2022) and Gans et al. (2019) emphasize how decision-makers design experiments to gather information that improves subsequent investments, product designs, or market strategies. Camuffo et al. (2024) examine how scientific training can help entrepreneurs design better experiments. In these studies, experimentation reduces uncertainty, leading to better decision-making. Our work shifts the focus to settings where experiments yield direct private payoffs but have broader market-level consequences. In our setup, once one experiment succeeds, additional successes do not increase the overall surplus because the total value is shared among all successful firms. This market feature creates a gap between what benefits individual firms and what is jointly beneficial. Moreover, this gap depends on the composition of the experimenters. Thus, we extend the literature to explore market-level technological diversity and innovation.

A key implication of our payoff structure is the business-stealing motive: if only one success matters, the most promising approach becomes the dominant individual choice, even though doing so reduces the expected payoff of others following the same approach. This creates rational herding, driven not by information cascades but by rent capture. Our model formalizes how this incentive interacts with correlated failure risks to reduce technological diversity, which aligns with Dasgupta and Maskin (1987), who demonstrate that when firms ignore the effect of their investments on others, their research portfolios will be too similar to each other.

We also build on the theoretical literature examining how firms choose experimental strategies. Our model shares similarities with Nelson (1961), and the subsequent literature in firm-level incentives to innovate, particularly how private incentives can lead to herding and clustering in specific technological trajectories (Lieberman and Asaba, 2006; Krieger, 2021). Product Development, such as Ding and Eliashberg (2002) and Girotra et al. (2007).

However, while Nelson assumes a single attempt per approach, we do not impose such constraints, allowing for a more flexible representation of experimentation under uncertainty. Bryan and Lemus (2017) also explore choices among competing innovation approaches. Some approaches offer more potential for follow-on research but may be harder. In their model, two features lead to excessive diversification, rather than herding: Firms race to be first, and firms do not capture the full value of their investment. However, by not distinguishing between approaches and experiments, they do not consider the case where a firm may mount two experiments, nor do they allow outcomes within an approach to be correlated.

The second domain to which we connect is the literature on herding and information cascades when firms learn from rivals (Lieberman and Asaba, 2006; Krieger, 2021). Formal models of information cascades are developed in Bikhchandani et al. (1992) and Banerjee (1992), where agents disregard private information and converge on specific actions based on observed behavior of others. Hoelzemann et al. (2022) examine the case where public information lead researchers to herd into a mediocre but known approach instead of exploring potentially more promising but also more uncertain approaches. In contrast, our framework focuses on static beliefs, where herding arises not from updating beliefs but from private incentives to pursue the most promising approaches. Excessive herding arises because firms ignore that projects that share an approach have correlated outcomes, and only one success matters for the joint payoff.

Finally, our work speaks to the relationship between firm size and the diversity of approaches. Larger firms, which are more likely to be able to conduct multiple experiments simultaneously, face different trade-offs than smaller firms limited to a single experiment. Cohen and Klepper (1992) argue that concentrating experiments within a few firms reduces the diversity of approaches. Our model diverges from this stream of research by considering non-additive payoffs, where the incremental value of success diminishes significantly with each additional success. This non-additivity, captured by our assumption that total private value is divided among all successful firms, affects diversity and innovation at the market

level.

3 Experiments and Approaches

3.1 What is an approach?

For any technical problem where a solution is not yet known, there may be multiple approaches to solve it. Each approach represents a theory about the underlying causal structure that explains outcomes as a function of a set of inputs (Camuffo et al., 2020; Sorenson, 2024). Consider the challenge of building a quantum computer. A key decision is how to physically realize qubits, the basic units of quantum information. Different firms have adopted fundamentally distinct approaches to this problem. For example, IonQ creates physical qubits using trapped-ion technology.[2] In contrast, Microsoft uses topological qubits, which rely on quasiparticles—collective excitations within a material—whose existence as yet awaits conclusive experimental basis.[3] Google's approach uses synthetic qubits, encoding and manipulating quantum information in superconducting circuits cooled to extremely low temperatures.[4] The choice of approach represents a significant strategic bet. That is, among innovative companies working to solve a given problem—whether it be curing a disease, developing generative AI algorithms, or developing self-driving cars—the choice of approach is one of the most critical decisions they make.

3.2 Implementing an approach

While firms may adopt the same approach, their success often hinges on how effectively they implement it. For example, in the case of quantum computing, Google is not the only company using superconducting qubits. Other firms, such as IBM, Rigetti, and Intel, have also

[2]For more details, see `https://ionq.com/company`

[3]For more details, see
`https://azure.microsoft.com/en-us/solutions/quantum-computing/technology/`

[4]For more details, see `https://quantumai.google/quantumcomputer`

chosen this approach but differ in their implementations of superconducting circuits. Even if superconducting circuits prove to be a viable pathway for building quantum computers, not all firms pursuing this approach may succeed. Their success will ultimately depend on the effectiveness of their specific implementations.

3.3 Two sources of uncertainty

The distinction between approaches and implementations highlights two sources of uncertainty in any experiment, both of which matter for the outcome. The first is *approach uncertainty*—whether the underlying hypothesis behind the approach is valid and capable of leading to an effective solution. The second is *implementation uncertainty*—whether the firm can successfully execute the necessary steps, such as designing a superconducting circuit or synthesizing a lead compound to drug a target. Figure 1 maps experimental outcomes along two dimensions: whether the approach is correct or wrong, and whether the implementation is successful or unsuccessful.

[Figure 1 about here.]

The upper-left quadrant represents the ideal scenario, where the approach is valid and the specific implementation is done well. Examples of such experiments include the use of CRISPR technology to edit genes. In the top-right quadrant, the approach is valid, but a specific implementation fails. An example of this is the Wright brothers' early attempts to invent the airplane. While their concept for heavier-than-air flight was correct, several implementations were unsuccessful. The bottom row highlights cases where the approach itself is incorrect, i.e., a dead end. The bottom-left quadrant is an interesting case where, despite a dead-end approach, an experiment may succeed, albeit not for solving the original problem. Serendipitous discoveries, such as that of penicillin by Alexander Fleming, belong in this category. Another example is the "audion", invented by Lee DeForest, which was originally developed as a radio wave detector, but its true value as an amplifier was recognized

by others rather than by DeForest. Lastly, in the lower-right quadrant, the approach is wrong, and the subsequent implementation efforts also fail. For instance, early tungsten filament lightbulbs were plagued by blackening on the inside of the bulb, and the proposed approach was to improve the vacuum inside the bulb. Implementation efforts failed, and it later turned out that the approach was also incorrect. Irving Langmuir suggested filling the bulb with an inert gas instead. He correctly hypothesized that oxidation of the filament was not to blame. Instead, the heated filament itself was emitting electrons that were deposited on the glass surface, blackening it. Inert gases inside the bulb scattered the electrons, solving the problem.

3.4 Case Study: Experimentation in Alzheimer's Disease

The distinction between approaches and implementation underscores an important implication: outcomes of experiments relying on the same approach are correlated. This connection highlights a critical challenge for innovation when many firms converge on a single hypothesis or theory. A prominent example is drug development in Alzheimer's Disease (AD). While numerous hypotheses exist about the causes of AD and potential ways to prevent its onset, an effective cure remains elusive. Compounding this difficulty is the considerable crowding around only a few approaches. Table 1 reports the number of AD drug development projects initiated by leading hypotheses between 1998 and 2008. Nearly 40% of projects focused on the beta-amyloid hypothesis, and almost 60% focused on just two hypotheses: beta-amyloid and cholinergic.

Although there have been three FDA-approved drugs targeting beta-amyloids, their benefits are very modest.[5] These drugs are known to cause serious side effects, such as brain bleeds, and have minimal impact on slowing the progression of AD in very early-onset patients.[6] Furthermore, recent research has raised doubts about the legitimacy of key beta-

[5]One of these drugs, Aducanumab, was discontinued in late 2024 because " many doctors and researchers said wasn't fully proven to work." https://www.wsj.com/tech/biotech/biogen-ends-aduhelm-program-in-shift-of-alzheimers-resources-163c897a

[6]https://www.wsj.com/articles/new-alzheimers-drug-shows-positive-results-but-side-

amyloid science (Piller, 2024, 2022). If the beta-amyloid hypothesis turns out to be wrong, the effectiveness of a firm's implementation becomes irrelevant: *all drug candidates relying on beta-amyloid will likely fail.*

[Table 1 about here.]

Which firms are crowding into the beta-amyloid and cholinergic hypotheses? Figure 2 examines projects initiated between 2007 and 2008, comparing Pfizer—which launched six distinct AD drug development projects during this period—to the 18 firms that each initiated only one project in the same timeframe. Only 2 out of Pfizer's 6 projects (33%) focused on beta-amyloid drugs. In contrast, among the single-experiment firms, 8 out of 18 projects (44%) targeted beta-amyloids. This is an example of a large experimenter diversifying across approaches and reducing the risk of correlated failures, whereas smaller experimenters—who make their decisions independently—crowd into popular approaches, increasing systemic vulnerability to approach-level failure.

[Figure 2 about here.]

4 Model

4.1 A model of market structure and approach diversity

In this section, we develop a simple framework to examine how firms choose between approaches to solve innovation problems. The model analyzes how market structure—the presence of many small versus few large experimenters—affects the diversity of approaches pursued and market-level outcomes. Our model has four key elements. First, we distinguish between approaches and implementations. Second, even the most promising approach is not certain to be viable, i.e., there is a non-zero probability that it could be a dead-end. By the same token, more than one approach could be viable. Third, we assume that the combined

effects-11669766449

payoff to all successful experiments is fixed and divided between all successful firms. In particular, a successful experiment yields a positive private payoff, but this payoff diminishes in proportion to the number of successful firms. Furthermore, a second success for the same firm has no incremental value. Finally, beliefs are static, and there is no learning.

4.2 Approaches versus implementations

There are two potential approaches $\{a, b\}$ from which a firm can choose to solve a problem. Let π_a and π_b represent the known probabilities that approach a and b, respectively, are viable solutions to the problem. We represent the probability that the implementation of approaches $\{a, b\}$ is effective with $\{p_a, p_b\}$. An experiment in approach a has a probability of success of $p_a\pi_a$, and the probability of success with approach b is $p_b\pi_b$. There is a common belief that a is the more promising approach so that $\pi_a p_a > \pi_b p_b$.

Consider two experiments, one in each approach. Let $x_a = 1$ if the experiment in a succeeds and $x_b = 1$ if the experiment in b succeeds. Since the approaches are independent (the viability of approach a is not related to the viability of approach b), then $Cov(x_a, x_b) = \pi_a\pi_b ab - (\pi_a a)(\pi_b b) = 0$. Now consider two experiments, both using approach a. Because the experiments use the same approach, their expected outcomes are correlated, since $Cov(x_a, x_b) = \pi_a a^2 - (\pi_a a)(\pi_a a) = a^2\pi_a(1 - \pi_a) > 0$.

Let v be the value from a successful experiment and c be the cost. All firms derive the same value from an experiment and face the same cost of experimentation.[7] The incremental benefit from the second success is zero (Dasgupta and Maskin, 1987). Thus, for a firm with two experiments, two successes yield the same payoff as one success. If two independent firms succeed, they split the payoff.

[7] Allowing some firms to appropriate higher rents (Cohen and Klepper, 1996a) affects the number of experiments (i.e., the decision to enter) but not the choice of approach.

4.3 Single and Multi-Experiment Firms and Market-Level Approach Diversity

If firm i uses approach a, its expected payoff is $vp_a\pi_a - c$, and if it uses approach b, its expected payoff is $vp_b\pi_b - c$. Given our assumptions, the firm should choose approach a for their experiment if $vp_a\pi_a - c \geq 0$, and otherwise, not enter at all. We now analyze the choice of approach for the next experiment under two scenarios: first, where a separate firm j considers the second experiment, and second, where firm i itself conducts the second experiment.

Case 1 – Two single-experiment firms: Firm j observes firm i's decision to use approach a, but firm i's experiment is ongoing, so outcomes are unknown. Firm j's expected payoff from also using approach a is $v\pi_a\left[p_a(1-p_a) + \frac{1}{2}p_a^2\right] - c$. [8] Note that both probabilities depend on π_a, the probability that approach a is viable. If firm j instead decides to use approach b, its expected payoff is $v\left[\pi_b p_b(1-\pi_a p_a) + \frac{1}{2}\pi_a\pi_b p_a p_b\right] - c$. The difference between these expected payoffs, by $\Delta_S = v\left[\pi_a p_a\left(1-\frac{1}{2}p_a\right) - \pi_b p_b\left(1-\frac{1}{2}\pi_a p_a\right)\right]$, where Δ_s represents the single-experiment firm's incentive to 'herd' into approach a.

Case 2 – One multi-experiment firm: Suppose firm j is not present. Instead, suppose firm i is considering a second experiment. As before, we assume the outcome of the first experiment is unknown. If firm i chooses to conduct its second experiment in approach a also, its expected payoff becomes $v\pi_a\left(1-(1-p_a)^2\right) - 2c$. Since firm i already has an experiment in approach a and there is no additional benefit to a second success, this payoff represents one minus the probability of failing in both experiments, minus the cost of two experiments. If firm i instead conducts its second experiment in approach b, its expected payoff is $v\left[\pi_a p_a(1-\pi_b p_b) + \pi_b p_b(1-\pi_a p_a) + \pi_a\pi_b p_a p_b\right] - 2c$. Taking the difference between these expected payoffs, we can derive an expression for firm i's incentive to use approach a

[8] Intuitively, firm j's payoff is the probability that only its experiment is successful, plus half the probability that both firms are successful, in which case expected payoffs will be shared equally. Alternatively, each firm has an even chance of being the winner which gets the entire payoff.

in its second experiment: $\Delta_M = v\left[\pi_a p_a(1-p_a) - \pi_b p_b(1-\pi_a p_a)\right]$[9]

Let Δ represent the difference between the multi-experiment firm's incentive Δ_M and the single-experiment firm's incentive Δ_S to crowd into the dominant approach.

$$\Delta = \Delta_M - \Delta_S = \frac{1}{2} v \pi_a p_a \left[\pi_b p_b - p_a\right] \leq 0 \iff \pi_b p_b \leq p_a \tag{1}$$

A multi-experiment firm is less likely to crowd into the dominant approach than two single-experiment firms when $p_a > \pi_b p_b$, which always holds given our assumption that $\pi_a p_a > \pi_b p_b$ (approach a is more promising). That is, a single firm conducting two experiments is more likely to try different approaches than two separate firms.[10] This occurs because a multi-experiment firm maximizes its chance of at least one success, reducing correlation in outcomes by trading off success probability per attempt.

Proposition 1 *Markets with a higher share of single-experiment firms will feature a lower diversity of approaches.*

Experiments using the most promising approach are more likely to succeed than experiments in less promising approaches. It follows that a market with greater diversity in approaches will have a smaller share of *successful* experiments.[11] In terms of the model above, if only approach A is chosen, each experiment has a probability of success of $\pi_a p_a$, whereas if both approaches are featured, it is $\frac{1}{2}(\pi_a p_a + \pi_b p_b) < \pi_a p_a$. In other words, markets with a greater diversity of approaches would be associated with lower share of successful experiments.

Proposition 2 *Markets with a lower diversity of approaches will have a higher share of successful experiments.*

[9] One could interpret investigating approach b as exploration whereas a second experiment in a could be interpreted as exploitation.

[10] In Appendix B, we generalize to multiple firms and show that the results continue to hold. We also show that when competition between two successful firms dissipates rents, firms will switch to choosing the less promising approach sooner. That is, the difference in incentives to diversify between single and multi-experiment firms will shrink when the total private value of innovation shrinks with the number of successful innovators.

[11] Implicitly, this assumes that firms are more likely to herd into more, rather than less, promising approaches.

By choosing varied approaches, large-scale experimenters include less promising approaches than small-scale experimenters. Therefore, markets with a higher share of small-scale experimenters will have a higher share of successful experiments. However, what matters for market outcomes is maximizing the probability of at least one successful experiment. Diversity increases this probability as long as the gap between the approaches is not great. Using the notation developed, a market with both experiments in a single approach will have at least one success with probability $\pi_a\left(1-(1-p_a)^2\right)$, whereas a market featuring one experiment in each approach will have at least one success with probability $[\pi_a p_a(1-\pi_b p_b)+\pi_b p_b(1-\pi_a p_a)+\pi_a\pi_b p_a p_b]$. The difference is $[\pi_a p_a(1-p_a)-\pi_b p_b(1-\pi_a p_a)]$, which is proportional to Δ_M. Therefore, diversity will be associated with a lower probability of complete failure as long as $\Delta_M \geq 0$ i.e., as long as a multi-experiment firm would choose a diverse set of approaches.

A multi-experiment firm will diversify only when that maximizes this probability, i.e., when $\Delta_M \leq 0$. Yet Δ_S, is strictly greater than Δ_M. Thus, for some parameter values, we have $\Delta_S > 0 > \Delta_M$. In this range, two single-experiment firms would both use the dominant approach, while one firm conducting both experiments would use different approaches. Critically, the probability of at least one success would be *lower* with two single-experiment firms.[12] Outside this range, the approach choices would be identical. Simply put, a multi-experiment firm would choose two distinct approaches when this maximizes the probability of at least one success. More formally, the market-level probability of at least one success is maximized with two distinct approaches if and only if:

$$\Delta_M = v\left[\pi_a p_a(1-p_a)-\pi_b p_b(1-\pi_a p_a)\right] \leq 0 \iff \frac{\pi_b p_b}{\pi_a p_a} > \frac{1-p_a}{1-\pi_a p_a} \tag{2}$$

Because $\pi_a p_a > \pi_b p_b$, a *necessary* condition for approach diversity with a multi-experiment firm is that $\pi_a < 1$, as otherwise the inequality in equation 2's second line cannot hold. When $p_a = 1$, the inequality must hold, providing a sufficient condition. That is, uncertain ap-

[12]Consider $\pi_a = 0.5, \pi_b = 0.4, p_a = 0.5, p_b = 0.5$. The probability of at least one success with two single-experiment firms is 0.375, whereas with a single firm conducting both experiments, it is 0.4.

proaches combined with effective implementation increase the likelihood that markets with diverse approaches yield a higher probability of at least one success.

Proposition 3 *Markets with greater diversity of approaches have a higher probability of at least one success.*

[Table 2 about here.]

Figure 3 connects the three propositions in a directed acyclic graph (DAG). Here we relax terminology to implicitly assume N firms, consistent with the empirical analysis which follows in Section 6. Moving from left to right, Proposition 1 is consistent with the first set of nodes: a greater average scale of experimenters in a market—whereby firms, on average, pursue multiple experiments—is positively associated with the diversity approaches pursued. Next, Proposition 2 implies that a lower diversity of approaches pursued—explained by a greater number of *small* scale experimenters—will be associated with higher average success. On the other hand, Proposition 3 states that a greater approach diversity is associated with a higher probability of at least one success.

[Figure 3 about here.]

4.4 Simulating market equilibria

Generalizing the two-firm, two experiments, setup makes analytical characterization intractable. We use numerical simulations to evaluate the robustness of the model's comparative statics. The simulation logic is described in detail in Appendix C. In the baseline setting, multiple firms choose sequentially whether and where to experiment, and the effect of prior entry on expected rents. Firms continue to enter as long as expected payoffs are positive. We exogenously vary the share of firms able to run two experiments and measure resulting market outcomes.

Figure C.1 illustrates these simulation results. Markets with a greater share of multi-experiment firms explore a greater diversity of approaches. In turn, this reduces the risk

that all experiments fail, even thought the probability that any single experiment succeeds falls. These patterns hold even when we allow for the discovery of new approaches (Section C.2).

Lastly, Appendix C.3 reports simulations under a full market-equilibrium implementation of the model. We choose parameter values to match sample moments on share of multi-experiment firms, total number of firms, and trials. Three results emerge. First, the long-run share of experiments using each approach converges to the ratio of the probability that the approach is viable, confirming that approach selection aggregates correctly in equilibrium. Second, the equilibrium distribution of experimenters allows us to recover the latent share of multi-experiment firms. Third, the same comparative statics derived analytically persist in equilibrium: markets with more multi-experimenters have a greater diversity of approaches, exhibit lower average success, and are less likely to experience total failure. Taken together, the simulations demonstrate that the core intuition of the model extends to more complex market settings that more closely align with the empirical context we study.

5 Data & Methods

5.1 Empirical Context: Early-stage drug development

The ideal empirical setting for us is one where there are a variety of possible approaches to a complex, technical problem, and firms enter with one or more experiments. One key requirement is that experimentation must be expensive. If experiments were cheap, all firms would naturally pursue multiple approaches, as in settings like software development à la Koning et al. (2022). Another desirable feature is that more successes do not create more private value in the aggregate and that competition among successful firms does not fully dissipate the total private value either. Therefore, we want to focus on situations where the incremental value of an additional success is small in the aggregate, and where successful firms can differentiate sufficiently to limit rent dissipation through competition.

We use early-stage drug development as the empirical testbed for our model. The development of novel therapeutic drugs is a technologically intensive, high-cost industry where, in many cases, one drug captures a significant majority of the market in a disease area.[13] The potential reward from success is substantial, but so is the cost of experimentation (DiMasi et al., 2016).

The drug development process can take up to 15 years (Hughes et al., 2011) and consists of three main stages: discovery, pre-clinical research, and human clinical trials (Frankel et al., 2023). Data on the discovery process are typically unavailable to archival researchers since these contain secret developments that firms do not disclose (Krieger et al., 2022). However, progression to pre-clinical research is increasingly reported. The choice to move a project into pre-clinical testing is a good approximation for the conditions of our model, namely it involves an important choice about what biological target to drug, which is largely informed by scientific knowledge in the public domain (Knowles and Gromo, 2003).[14]

The first data source we use is Pharmaprojects, a commercial database curated by Citeline. Pharmaprojects provides comprehensive global drug research and development history. Drugs are tracked from early pre-clinical development through market launch and are retained regardless of their outcome. New drugs and events for existing drugs are added to the database daily after a thorough editorial process governed by industry experts who scour thousands of sources in the public domain (e.g., press releases).

We collect data from the Pharmaprojects Trends database, which provides annual updates for the status of drug development projects since 1995.[15] A status corresponds to a clinical trial phase or the drug being discontinued or launched. It is important to distinguish between a lead compound and a drug development project. For our purposes, a drug development project concerns the clinical trials (development) for a lead compound intended

[13]This is true at least during the first few years that the drug is on the market and still has patent protection.

[14]Evaluating the promise of a target has also benefited from the recent surge in the availability of big data (Chen and Butte, 2016) and machine learning algorithms (Dara et al., 2022).

[15]These data were downloaded in August 2024.

for a particular therapeutic class.[16] We highlight this distinction as it is common for a lead compound to be in clinical trials for more than one therapeutic application. The raw data contain 46,390 distinct lead drug compounds, which correspond to 98,784 drug development projects (i.e., a lead compound is, on average, in development for two therapeutic indications). We keep projects where (i) we can identify the year in which pre-clinical trials began,[17] (ii) the organization leading the project is a for-profit firm, and (iii) projects where we can identify the biological target that the drug is intended to act upon. These data also allow us to observe important outcomes for each project, namely, if the project advances to phase 1 clinical trials and if the project results in a drug launch. After these cleaning steps, we have 49,866 drug development projects started by 3,845 firms over 28 years between 1996 and 2023 across 241 therapeutic classes.

We also collect data from two additional sources. First, from *Pitchbook* we collect data on firm founding years and ownership status. We use a fuzzy matching algorithm to match Pharmaprojects to Pitchbook by firm name (see Appendix D.2 for an explanation of the match and its accuracy). Second, we collect data from the GWAS catalog, a record of all scientific publications in top-tier journals reporting the discovery of new target-disease correspondences (Tranchero, 2023). We use these data to measure the rate of new target discovery across therapeutic classes over time.

5.2 Construction of measures

Approaches and implementation: We define an approach as the choice of therapeutic target (Thomke et al., 1998; Tranchero, 2023), and an implementation is a specific lead compound for that target in a preclinical trial. An experiment consists of choosing a target and a candidate molecule to "drug the target" in a pre-clinical trial.

[16] In Pharmaprojects, a therapeutic class most closely resembles the Anatomical Therapeutic Chemical (ATC) Classification Level 3, which represents chemical, pharmacological, or therapeutic subgroups.

[17] any projects that first appear in the data in 1995 because we lack information on whether projects that first appear in 1995 also began in that year. We also drop projects that began in 2024 as the data are incomplete for this year.

Our unit of analysis is the market, defined as a therapeutic class-year (i, t).[18] We define market observations within a year, as this will be most consistent with our theoretical assumption of no updating of beliefs within an observation period. In other words, we assume that a firm that starts two projects in a therapeutic class–year does not know the outcome of the first project before starting the second. In Appendix E, we replicate our baseline results for two-year and 5-year windows and find qualitatively similar trends. We only include therapeutic class-years with (i) at least two firms and (ii) at least two approaches, because when only one target is in use, there is, by construction, no diversity in approaches.

Tables 3 and 4 present descriptive statistics and the correlation matrix for the variables in our market-level dataset. The dataset comprises 2,523 observations at the therapeutic class-year (market) level, spanning 139 therapeutic classes. On average, each market features 9.8 distinct approaches deployed across 11.6 experiments, 13.8% of the experiments are initiated by multi-experiment firms, and the average firm conducts 1.2 experiments.

[Table 3 about here.]

[Table 4 about here.]

Market Diversity of Approaches. We measure the market-level diversity of approaches using Shannon entropy (Shannon, 1948), a popular measure used in ecology to quantify the abundance and evenness of species present in a community (Margalef, 1958).[19] Formally, let p_k be the proportion of projects using target k for all drug development projects in disease i started in year t, *Target Diversity* is

$$Target\ Diversity_{i,t} = -\sum_{k=1}^{n} p_k \ln p_k$$

[18] We collapse the project-level data to the therapeutic class-year level and then construct market-level measures of interest. Our model, though based on individual firm-level behavior, addresses market-level rather than firm-level outcomes. Diversity at the individual firm level has little meaning for single-experiment firms. More substantively, what matters is whether at least one of the experiments launched in the market succeeds.

[19] Recent studies of the science of science also use Shannon entropy to quantify the certainty of knowledge (Kang et al., 2024). Using the Herfindahl–Hirschman index gives very similar results, as shown in Appendix Table E.2.

Success. We define the success of a drug development project as whether the project advances from pre-clinical experimentation to a phase 1 clinical trial. Progression to phase 1 clinical trials can be interpreted as a signal of *technical* success, since a project will only progress to human trials if the preclinical trial provides evidence of safety and efficacy. We create two measures based on this measure of success. The first, *At least 1 Pre-Clinical Success*, is a dummy indicator equal to 1 if in market-year (i, t) at least one project progresses to phase 1 trials. The second variable, labelled *Share of Pre-Clinical Success*, is the *share* of projects that progress to phase 1 clinical trials..

In additional analysis (see Appendix E), we also measure success as the ultimate launch of a drug. While drug launch also signals technical success insofar as the drug has been proven safe and effective in humans, it also represents *commercial* considerations. The cost of bringing a drug rises steeply through the latter stages of human clinical trials. Hence, a firm's decision to proceed with later stages also depends on economic factors such as drugs in its development pipeline and expected competition from drugs by other firms in the market.

Experimenter Scale. The main independent variable in our analysis captures the distribution of experimenter scale within a market. Our preferred measure, *Average Experimenter Scale*, is defined as the average number of drug development projects initiated (i.e., beginning preclinical trials) by firms in therapeutic class j in year t.[20]

Discovery. The variable *ln(Discovery)* is the natural logarithm of the count of publications in GWAS in year t that are the first to report a new relationship between a target and therapeutic class i. The GWAS catalog only reports the disease and/or trait that a particular study addresses, so to match these data to Pharmaprojects, we create a correspondence between diseases/traits and therapeutic classes (the details of this correspondence are explained in Appendix D).

[20] We also create the variable *Multi-Experiment Share*, which represents the share of firms in market-year (i, t) that initiate preclinical trials for two or more distinct projects. These measures are highly correlated, with a Pearson correlation coefficient of 0.82. Thus, we do not report results with *Multi-Experiment Share* as they are similar to specifications that use *Average Experimenter Scale*. See also Appendix A for a discussion of measurement error and its implications for our results.

Control Variables. In all specifications, we include *ATC-1*× *Year* fixed effects (Branstetter et al., 2022). ATC-1 represents the first level in the anatomical therapeutic class classification and most closely resembles distinct industry classifications, such as Neurology, Cardiovascular, and Dermatologicals. This interacted fixed effect thus captures time-varying differences in industry-specific characteristics, such as technology or demand shocks. Some specifications also include a set of market structure controls, which in all tables are denoted as *Market Structure Controls*. This includes the number of firms, targets, and projects started for each therapeutic class–year.[21]

6 Results

Lacking a source of exogenous variation in experimenter scale, what we document empirically are associations. First, we provide correlational evidence consistent with the three propositions derived from our model. Next, we discuss alternative explanations and explore the extent to which they can explain the empirical patterns. Lastly, we explore which firms are more likely to launch multiple experiments and which firms try novel approaches.

Establishing causality in this context is challenging. In particular, identifying a causal link running from experimenter scale to diversity, and then to outcomes, requires variation in experimenter scale that is not confounded with factors that affect diversity or outcomes. The ideal econometric experiment would involve identifying a source of exogenous variation in the cost or value of experimentation for certain firms in a market. For instance, a reduction in the cost of experimentation for some firms within a market could lead to an increase in the number of experiments they initiate, thereby altering the average experimenter scale without directly affecting the diversity of approaches, and also not directly affecting the success probability. Shocks that influence entire markets —for example, the introduction of

[21] We recognize that these are themselves determined in equilibrium by underlying fundamentals such as the attractiveness of the market, the available approaches, and the cost of entry. Our empirical analysis intends to check consistency with theory rather than identify causal relationships. Our results are not sensitive to their inclusion.

Medicare Part D in 2004, which arguably increased the value of drugs for elderly patients (Dranove et al., 2022)— tend to affect *all* firms operating in the market. As such, these shocks would increase the number of experiments but also the number of entrants and the effect on average experimenter scale is ambiguous. Moreover, they may also affect diversity directly by encouraging the exploration of hitherto unexplored approaches.

6.1 Diversity of Approaches and the Success of Experimentation

We begin by examining the relationships theorized by Propositions 1 and 2 between experimenter scale, the diversity of approaches, and the average success of experimentation using the following specification:

$$y_{i,t} = \alpha + \beta_1 x_{i,t} + \theta_{i,t} + \gamma_i \times \tau_t + \epsilon_{i,t} \tag{3}$$

In this specification $y_{i,t}$ represents either *Target Diversity*, or *Share of Pre-Clinical Success*, and $x_{i,t}$ is the average experimenter scale in a therapeutic class–year. We estimate this specification with and without $\theta_{i,t}$, a set of market structure controls: the number of firms, targets, and pre-clinical experiments within a therapeutic class–year. In all models, we include ATC-1×Year fixed effects $\gamma_i \times \tau_t$, and standard errors are clustered at the ATC-1 level to account for correlations in errors among market observations within the same industry group.

[Figure 4 about here.]

6.1.1 Experimenter Scale and Approach Diversity

Figure 4 shows the non-parametric relationships, which closely mirror the parametric relationships reported in Tables 5 and 6. Table 5 column (1) reports the baseline relationship between experimenter scale and target diversity, showing a positive and significant association. In column (2), we add ATC-1×Year fixed-effects, and in column (3), we incorporate

market structure controls. Notably, the estimated coefficient for β_1 remains consistent across models. Based on the estimate in column (3), our findings suggest that a one-unit increase in the average experimenter scale (e.g., from an average of 1.1 to 2.1 experiments per firm) is associated with a 0.651-unit increase in target diversity, which corresponds to approximately a 0.76 standard deviation increase in the diversity of targets deployed in a therapeutic class–year. This result is consistent with Proposition 1.

[Table 5 about here.]

6.1.2 Approach Diversity and Experiment Success

Our theoretical framework predicts that experimenter scale shapes the diversity of approaches pursued in a market, and that this diversity, in turn, affects innovation outcomes (see the DAG in Figure 3). To test this, we first estimate how *Target Diversity* varies with *Average Experimenter Scale*, then examine how diversity relates to the success of experimentation.

Average Experimenter Success Table 6 column (1) presents an OLS estimate of the relationship between diversity and the share of pre-clinical successes, which is small and statistically insignificant. In contrast, the 2SLS estimate in column (2), which uses the component of *Target Diversity* predicted by *Average Experimenter Scale*, reveals a negative and statistically significant effect. Specifically, a one standard deviation increase in target diversity is associated with a 3.9 percentage point decrease in the average success rate of experiments.[22]

This difference between the OLS and 2SLS estimates is informative. The OLS estimate reflects the combined effect of different sources of diversity, some associated with less promising approaches, others with genuinely valuable discoveries. In contrast, the 2SLS estimate

[22]Though we use the term "instrument" and a 2SLS setup, the objective is primarily to test the mechanism linking experimenter scale, diversity, and success, rather than deal with the potential endogeneity of the experimenter scale itself.

isolates the component of diversity due to experimenter scale, which our model predicts should be negatively associated with average success. That is, the 2SLS approach helps separate the component of diversity that arises from broader experimentation efforts from the component driven by the discovery of promising new targets. The former is associated with lower average success, whereas the latter arguably is associated with higher average success.

[Figure 5 about here.]

The DAG presented in Figure 5 illustrates the underlying intuition. The discovery of new targets—our proxy for an influx of novel approaches—is likely to increase the diversity of targets observed in a market. At the same time, if newly discovered targets are also more scientifically promising, discovery may have a direct positive effect on experimental success. This dual role of discovery implies that *Target Diversity* captures two components: one driven by experimenter scale, and another reflecting the emergence of promising new approaches.

At Least One Success This distinction becomes especially relevant when considering Proposition 3, which focuses on the probability that a market achieves at least one successful experiment. In this case, both components of diversity—scale-driven exploration and new target discovery—should positively contribute to the likelihood of at least one success. As a result, the OLS estimate reflects the combined positive effect of both sources. However, the 2SLS estimate, which uses the component of *Target Diversity* predicted by *Average Experimenter Scale*, removes the component attributable to new target discovery, leaving only the scale-driven portion. Therefore, we would expect the 2SLS estimate to be smaller than the OLS estimate.

Column (4) of Table 5 presents the OLS result, showing that a one standard deviation increase in Target Diversity is associated with a 26.8 percentage point increase in the probability of at least one pre-clinical success. When using the component of diversity predicted by experimenter scale, the resulting 2SLS estimate is slightly smaller (column (5)), though

the difference is not statistically significant. These patterns are consistent with the predictions of our model. In the next section, we analyse alternative explanations for the observed empirical patterns.

[Table 6 about here.]

6.2 Alternative Explanations

We reiterate that we do not estimate the causal effects of experimenter scale. Instead, our objective is to assess whether the relationships among scale, diversity, and outcomes are consistent with the logic of the model. Where we use instrumental variables, we do so as a way to explore underlying mechanisms and rule out certain interpretations, rather than to claim estimation of causal effects.

In this section, we take a multipronged approach to evaluate alternative explanations for our empirical results and the robustness of our results to our measurement choices.

6.2.1 Alternative theories and unobserved heterogeneity

Logical Consistency of Findings Alternative explanations, including likely sources of unobserved heterogeneity, are unlikely to explain the pattern of empirical results regarding the success of experimentation (Propositions 2 and 3): that greater experimenter scale increases the probability of at least one success but lowers average success, and these effects involve market level diversity. For instance, suppose that some markets have more technical opportunities (higher probability of experiments succeeding), and this is positively (negatively) associated with the number of approaches observed. Then, there should be a positive (negative) relationship between diversity and average success, but also a positive relationship with at least one success. Moreover, there would be no reason to expect a relationship between diversity and average scale of experimenters.

Alternatively, suppose experimenter scale is positively related to technological opportunity or superior firm capability, which, in turn, is positively related to the diversity of

approaches. However, this should also *increase*, not decrease, average success. Thus, while we cannot establish a causal link between experimenter scale and target diversity, the observed associations between experimenter scale, target diversity, and success outcomes are consistent with the hypothesized causal relationship. Table 7 summarizes the arguments.

[Table 7 about here.]

Discovery of new targets In Table 8 columns (1) through (4), we seek to explicitly control for the discovery of new targets. Unfortunately, in doing so, we lose about two-thirds of the sample, resulting in very imprecise estimates. Overall, the results are similar to those reported in Tables 5 and 6. The coefficient on *Average Experimenter Scale* decreases in magnitude when $\ln(Discovery)$ is included, and the relationship between *Target Diversity* and at least one success is largely unchanged when controlling for discovery.

[Table 8 about here.]

Firm heterogeneity Second, there may be firm characteristics within a market that are correlated with both the number of experiments firms choose to launch and their choice of approach. For instance, if younger firms are less likely to conduct multiple experiments but also more likely to pursue novel approaches, the average firm age in a market could be correlated with target diversity. Table 8 columns (5) through (8) similarly show that our results are qualitatively unchanged when we control for the average age of experimenters.

Approach uncertainty The next analysis addresses the complication that arises when a target is known to be a valid approach. In our model, this corresponds to an approach where π—-the probability that the approach is a viable solution to the problem—-equals one, meaning there is *no* uncertainty about its viability. When $\pi = 1$, the key implication is that outcomes for experiments using the same approach are no longer correlated, as success depends solely on implementation (see Appendix B). In this case, firms will select the

approach with the highest perceived probability of implementation success. To address this, we restrict the underlying project-level dataset to projects where the target chosen has not been previously successfully drugged, which we measure by the first launched product. We report these results in Table 9. Comparing column (2) from Table 9 with Table 5, we see a larger coefficient on *Average Experimenter Scale*, suggesting a downward bias on the original estimate due to the inclusion of target–therapeutic class combinations already known to be viable. In particular, the results in Table 9 imply that a one-unit increase in average experimenter scale is associated with a 0.93 standard deviation increase in target diversity. The coefficient estimate in column (4) of Table 9 on Target Diversity is greater in magnitude than in Table 6, which suggests our core results overestimate the magnitude of the negative relationship between target diversity and average success. In contrast, the results for at least one success in both tables are indistinguishable.

[Table 9 about here.]

6.2.2 Robustness to alternative measurement

Differing time windows We also test the sensitivity of our results to alternative time windows. Recall that the unit of analysis in our baseline results is the therapeutic class–year. The choice to look within a single year is arbitrary, so we construct two additional estimation samples, where we define observations within a two-year and five-year window. Table E.1 in the Appendix reports these results. Although we lose the number of observations through a higher level of aggregation, we find results consistent with our preferred specification, which defines observations within a 1-year period. The only exception is that we lose statistical significance for the relationship between average experimenter scale and average success in the therapeutic class–five-year period data, and find no relationship instead.

Alternative measures of diversity and success Table E.2 in the Appendix shows that our results are robust to alternative measures of key variables. In panel (a) our results are

similar if we measure target diversity with the Herfindahl-Hirschman Index. In panel (b) we define success as whether a pre-clinical experiment ultimately results in a drug product launch. Measuring success as drug launch are consistent with our Phase I findings reported in Table 6, albeit the results are noisier.[23]

6.3 Which firms do multiple experiments and introduce novelty?

Our data and results raise at least two additional questions, which we explore in this section. First, we have not addressed which firms choose to launch multiple experiments. One potential mechanism (as demonstrated in our simulations in Appendix C) involves differences across firms in the cost of experimentation. Firms with a lower cost of experimentation or, equivalently, with a greater ability to capture value, are more likely to launch multiple experiments.

A large body of research suggests that firm size likely explains this pattern. Specifically, multi-experiment firms are most likely *large*, as they benefit from lower experimentation costs (Cohen and Klepper, 1996b) and can appropriate greater value from experiments (Arora et al., 2023). This advantage stems, in part, from their complementary downstream co-specialized assets, which are essential for commercialization (e.g., Rosenbloom, 2000; Filippetti and D'Ippolito, 2017).

We use firm ownership as a proxy for firm size, distinguishing between private and public firms. Interestingly, the data provide no evidence that public firms are more likely to conduct multiple experiments in a market. In panel (A) of Figure 6, we show that private firms conduct only one experiment in a therapeutic class–year 61.5% of the time. For public firms, this share is slightly *higher*, at 63.3%.[24] Panel (B) presents the raw counts of experiments by experimenter scale. While public firms conduct significantly more experiments overall, most of the time, they launch only one pre-clinical experiment in a therapeutic class–year.

[23] While we believe that commencing Phase I trials is our best proxy of *technological* success, because later stage trials involve commercial considerations as well, phase I trials are distant from financial payoffs.

[24] These shares are likely higher than the true values, due to underreporting and data availability, but the relative shares ought not to be biased.

Second, the consideration of the discovery of new targets adds an intriguing nuance: which firms are more likely to introduce novelty? Specifically, when a new target–therapeutic class link is discovered, which firms—if any—are more likely to be the first to pursue this approach in pre-clinical trials? In panel (C) of Figure 6, we plot the share of experiments by single-experiment firms and multi-experiment firms (those conducting more than one experiment in a therapeutic class–year) that are the first to use a target in a therapeutic class.[25] The figure shows that single-experiment firms are more likely to pursue a novel target: 54.9% of experiments by these firms introduce a new target, compared to 45.9% for multi-experiment firms. However, panel (D) reveals no significant difference between private and public firms in their likelihood of pursuing a novel approach.

These results highlight an important trade-off. Markets dominated by small-scale experimenters are less favorable for innovation in some respects, as they are more likely to rely on a less diverse set of approaches and are less likely to achieve at least one successful experiment. However, these markets are also more likely to feature new targets.

[Figure 6 about here.]

7 Discussion

This paper addresses the question of how market structure influences the diversity of experimental approaches firms take and, ultimately, the likelihood of solving innovation problems with uncertain solutions. To investigate this, we develop a theoretical model that explores how multi-experiment and single-experiment firms choose between approaches. The model yields three propositions. First, firms conducting multiple experiments are more likely to diversify their approaches, whereas single-experiment firms tend to converge on the most promising approach, reducing market-level diversity. Second, markets dominated by single-

[25] To improve reliability, we include only projects initiated since 2005, as earlier projects are more likely to represent the first target–therapeutic class combinations due to the dataset's starting point in 1995, which excludes targets drugged before this year.

experiment firms achieve higher success rates per experiment, due to their focus on the most promising approach. Third, the probability of at least one success is higher in markets with a larger share of multi-experiment firms, due to the greater diversity of approaches. To test these predictions, we use detailed data on early-stage pharmaceutical drug development, spanning 49,866 projects across 27 years, and demonstrate how experimenter scale drives approach diversity and influences market-level success.

The findings reveal a robust relationship between experimenter scale and approach diversity. A one-unit increase in the average experimenter scale is associated with a 0.76 standard deviation increase in the diversity of targets deployed in a therapeutic class–year. Importantly, markets with more target diversity are more likely to achieve at least one successful outcome. A unit standard deviation increase in target diversity is associated with a 26.8 percentage point increase in the probability that a market has at least one pre-clinical success, on average. However, this benefit comes at the cost of a lower average success rate per experiment, as larger-scale experimentation dilutes focus on the most promising approaches. Indeed, the same increase in target diversity is associated with a 3.9 percentage point decrease in the share of projects that succeed in pre-clinical development. These results are inconsistent with alternative explanations, and robust to controls for new target discovery and firm age.

Our findings have implications for the broader literature on experimentation and innovation. By highlighting the dual roles of small- and large-scale experimenters, we add to the understanding of how diversity in approaches arises (Cohen and Klepper, 1992). Some small-scale experimenters act as pioneers, introducing novel approaches that expand the frontier of possibilities. However, there is another aspect to scale. Smaller firms are more likely toherd into the most promising approach. This pattern reflects a business-stealing motive in experimentation: private incentives push firms toward the most promising approach, ignoring that some of their expected payoffs come at the expense of others, particularly those exploring the same approach as them. Large-scale experimenters provide depth by thoroughly test-

ing and validating established approaches (Kotha et al., 2011; Ahuja and Morris Lampert, 2001). Our distinction between approach and implementation further informs studies of experimentation by underscoring the importance of considering the shared risks common to all experiments using the same approach. Failing to separate these components can obscure whether failure stems from flawed hypotheses or flawed execution.

Policy Implications. For policy, our results suggest a nuanced perspective on fostering innovation through experimentation. Encouraging both small- and large-scale experimenters within markets is crucial to achieving a balance between exploration and exploitation. Policies that support small-scale experimenters, such as grants for early-stage research or incubator programs, can help uncover novel targets and approaches (Bradley et al., 2021). Meanwhile, mechanisms that enable larger firms to scale their experimentation, such as tax incentives or public-private partnerships, ensure that promising but understudied approaches receive the rigorous testing needed for broader application (Dimos and Pugh, 2016; Lerner, 2009). Perhaps the most contentious implication is for anti-trust. Whereas tech giants and large firms are viewed as a source of reduction in diversity, our findings sound a note of caution. Insofar as market leaders are themselves engaged in research, they increase, not decrease, diversity of approaches. By designing interventions that balance the discovery of novel approaches and the more thorough exploration of existing ones, policymakers can enhance the robustness of innovation ecosystems and increase the likelihood of market-level breakthroughs.

These insights also have practical implications for the structure of research funding and the design of innovation ecosystems. Allocating resources to create spaces where small-scale experimenters can explore untested ideas while facilitating partnerships with larger organizations can improve innovation outcomes (Cappelen et al., 2012). For instance, collaborative frameworks that combine the agility of startups with the resources of established firms may foster a more diverse and productive experimental landscape (Polidoro Jr and Yang, 2021;

Bhaskaran and Krishnan, 2009). Similarly, strategies that mitigate herding, such as incentivizing exploration of high-risk, high-reward approaches, can prevent over-concentration on seemingly safe options, thus ensuring a more resilient innovation pipeline (Von Essen et al., 2020). This intuition also has implications for recommender systems (Fleder and Hosanagar, 2009) and the rapid diffusion of generative AI (Doshi and Hauser, 2024). These typically winnow the set of approaches being considered to the most promising ones—which is good for individual-level outcomes—but may come at the cost of overall market-level diversity.

Limitations and Directions for Future Research. While our findings provide valuable insights, several limitations warrant discussion. First, the causal interpretation of our results remains challenging. For example, simultaneous drivers of experimenter scale and diversity, like regulatory changes or scientific breakthroughs, complicate causal attribution. Although we use extensive controls, unobserved factors—such as market-specific shocks to funding, technology, or regulation—may affect our estimates. However, such factors are unlikely to account for the full set of empirical associations we document.

Second, our focus on the pharmaceutical industry, while offering rich data, limits generalizability. Pharmaceuticals involve high R&D costs, long timelines, and significant uncertainty, which may amplify the dynamics we observe. Other industries, with different innovation cycles or competitive pressures, might display distinct patterns. Exploring these relationships in contexts like materials or renewable energy could test the broader applicability of our model and findings. With regard to our theory, our assumptions imply two key scope conditions: (1) the total value from experimentation is fixed and independent from the *number* of successes, and (2) that there is variation across firms in their ability to conduct multiple experiments. At first, these scope conditions may seem restrictive, but we believe that there are many markets and technologies—such as batteries, quantum computing, and medical devices—where these conditions appear to be satisfied. In these industries, among others, experimentation is costly, which is one reason for variation in experimenter scale.

Furthermore, all of these examples share the promise of a dominant successful approach, such as lithium-ion in batteries, where the successful firm(s) will share temporary monopoly rents.

Third, while our conceptual framework and empirical analysis center on the diversity of approaches, novelty may also arise along other dimensions. A single-shot experiment may have the same target as others but still pursue a chemically differentiated solution. However, because success depends on the underlying approach being viable, these solutions share correlated technical risk. Future research could investigate how approach- and solution-level novelty interact.

By studying how market structure shapes collective experimentation outcomes, we hope this work encourages further analysis of how innovation systems can be designed to solve the complex and pressing problems.

References

Ahuja, G. and Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, 22(6-7):521–543.

Arora, A., Cohen, W., Lee, H., and Sebastian, D. (2023). Invention value, inventive capability and the large firm advantage. *Research Policy*, 52(1):104650.

Ayala-Fontanez, N., Soler, D., and McCormick, T. (2016). Current knowledge on psoriasis and autoimmune diseases. *Psoriasis: Targets and Therapy*.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817.

Bhaskaran, S. R. and Krishnan, V. (2009). Effort, revenue, and cost sharing mechanisms for collaborative new product development. *Management Science*, 55(7):1152–1169.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026.

Bradley, S. W., Kim, P. H., Klein, P. G., McMullen, J. S., and Wennberg, K. (2021). Policy for innovative entrepreneurship: Institutions, interventions, and societal challenges. *Strategic Entrepreneurship Journal*, 15(2):167–184.

Branstetter, L., Chatterjee, C., and Higgins, M. J. (2022). Generic competition and the incentives for early-stage pharmaceutical innovation. *Research Policy*, 51(10):104595.

Bryan, K. A. and Lemus, J. (2017). The direction of innovation. *Journal of Economic Theory*, 172:247–272.

Camuffo, A., Cordova, A., Gambardella, A., and Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, 66(2):564–586.

Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., and Spina, C. (2024). A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*.

Cappelen, Å., Raknerud, A., and Rybalka, M. (2012). The effects of R&D tax credits on patenting and innovations. *Research Policy*, 41(2):334–345.

Chen, B. and Butte, A. (2016). Leveraging big data to transform target selection and drug discovery. *Clinical Pharmacology & Therapeutics*, 99(3):285–297.

Ciarli, T., Kenney, M., Massini, S., and Piscitello, L. (2021). Digital technologies, innovation, and skills: Emerging trajectories and challenges. *Research Policy*, 50(7):104289.

Cohen, W. M. and Klepper, S. (1992). The tradeoff between firm size and diversity in the pursuit of technological progress. *Small Business Economics*, 4:1–14.

Cohen, W. M. and Klepper, S. (1996a). Firm size and the nature of innovation within industries: the case of process and product R&D. *The Review of Economics and Statistics*, pages 232–243.

Cohen, W. M. and Klepper, S. (1996b). A reprise of size and R & D. *The Economic Journal*, 106(437):925–951.

Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., and Ahsan, M. J. (2022). Machine learning in drug discovery: a review. *Artificial Intelligence Review*, 55(3):1947–1999.

Dasgupta, P. and Maskin, E. (1987). The simple economics of research portfolios. *The Economic Journal*, 97(387):581–595.

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of Health Economics*, 47:20–33.

Dimos, C. and Pugh, G. (2016). The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature. *Research Policy*, 45(4):797–815.

Ding, M. and Eliashberg, J. (2002). Structuring the new product development pipeline. *Management Science*, 48(3):343–363.

Doshi, A. R. and Hauser, O. P. (2024). Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290.

Dranove, D., Garthwaite, C., and Hermosilla, M. (2022). Does consumer demand pull scientifically novel drug innovation? *The RAND Journal of Economics*, 53(3):590–638.

Filippetti, A. and D'Ippolito, B. (2017). Appropriability of design innovation across organisational boundaries: exploring collaborative relationships between manufacturing firms and designers in italy. *Industry and Innovation*, 24(6):613–632.

Fleder, D. and Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5):697–712.

Frankel, A. P., Krieger, J. L., Li, D., and Papanikolaou, D. (2023). Evaluation and learning in R&D investment. *National Bureau of Economic Research*.

Gans, J. S., Stern, S., and Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5):736–756.

Girotra, K., Terwiesch, C., and Ulrich, K. T. (2007). Valuing r&d projects in a portfolio: Evidence from the pharmaceutical industry. *Management Science*, 53(9):1452–1466.

Gross, D. P. and Sampat, B. N. (2023). The World War II crisis innovation model: What was it, and where does it apply? *Research Policy*, 52(9):104845.

Guo, J., Zhang, H., Lin, W., Lu, L., Su, J., and Chen, X. (2023). Signaling pathways and targeted therapies for psoriasis. *Signal Transduction and Targeted Therapy*, 8(1):437.

Hoelzemann, J., Manso, G., Nagaraj, A., and Tranchero, M. (2022). The streetlight effect in data-driven exploration. *UC Berkeley and University of Vienna*.

Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249.

Kang, H., Fu, L., Funk, R. J., Wang, X., Ding, J., Liang, S., Wang, J., Zhou, L., and Zhou, C. (2024). Scientific and technological knowledge grows linearly over time. *arXiv preprint arXiv:2409.08349*.

Knowles, J. and Gromo, G. (2003). Target selection in drug discovery. *Nature Reviews Drug Discovery*, 2(1):63–69.

Koning, R., Hasan, S., and Chatterji, A. (2022). Experimentation and start-up performance: Evidence from A/B testing. *Management Science*, 68(9):6434–6453.

Kotha, R., Zheng, Y., and George, G. (2011). Entry into new niches: The effects of firm age and the expansion of technological capabilities on innovative output and impact. *Strategic Management Journal*, 32(9):1011–1024.

Krieger, J., Li, D., and Papanikolaou, D. (2022). Missing novelty in drug development. *The Review of Financial Studies*, 35(2):636–679.

Krieger, J. L. (2021). Trials and terminations: Learning from competitors' r&d failures. *Management Science*, 67(9):5525–5548.

Lerner, J. (2009). *Boulevard of broken dreams: why public efforts to boost entrepreneurship and venture capital have failed–and what to do about it*. Princeton University Press.

Lieberman, M. B. and Asaba, S. (2006). Why do firms imitate each other? *Academy of Management Review*, 31(2):366–385.

Liu, P.-P., Xie, Y., Meng, X.-Y., and Kang, J.-S. (2019). History and progress of hypotheses and clinical trials for alzheimer's disease. *Signal Transduction and Targeted Therapy*, 4(1):29.

Margalef, R. (1958). Information theory in ecology. *General Systems*, 3:36–71.

Nelson, A., Anthony, C., and Tripsas, M. (2023). "If I could turn back time": Occupational dynamics, technology trajectories, and the reemergence of the analog music synthesizer. *Administrative Science Quarterly*, 68(2):551–599.

Nelson, R. R. (1961). Uncertainty, learning, and the economics of parallel research and development efforts. *The review of economics and statistics*, pages 351–364.

Nestle, F. O., Kaplan, D. H., and Barker, J. (2009). Psoriasis. *New England Journal of Medicine*, 361(5):496–509.

Piller, C. (2022). Blots on a field? *Science (New York, NY)*, 377(6604):358–363.

Piller, C. (2024). Researchers plan to retract landmark alzheimer's paper containing doctored images. *Science*.

Polidoro Jr, F. and Yang, W. (2021). Corporate investment relationships and the search for innovations: An examination of startups' search shift toward incumbents. *Organization Science*, 32(4):909–939.

Rosenbloom, R. S. (2000). Leadership, capabilities, and technological change: The transformation of ncr in the electronic era. *Strategic Management Journal*, 21(10-11):1083–1103.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Sorenson, O. (2024). Theory, search, and learning. *Strategy Science*.

Tan, D. (2023). The road not taken: Technological uncertainty and the evaluation of innovations. *Organization Science*, 34(1):156–175.

Thomke, S., Von Hippel, E., and Franke, R. (1998). Modes of experimentation: an innovation process—and competitive—variable. *Research Policy*, 27(3):315–332.

Tranchero, M. (2023). Finding diamonds in the rough: Data-driven opportunities and pharmaceutical innovation. *Working Paper. University of California, Berkeley*.

Von Essen, E., Huysentruyt, M., and Miettinen, T. (2020). Exploration in teams and the encouragement effect: Theory and experimental evidence. *Management Science*, 66(12):5861–5885.

Figure 1: A 2x2 framework of experimentation: approaches vs. implementation

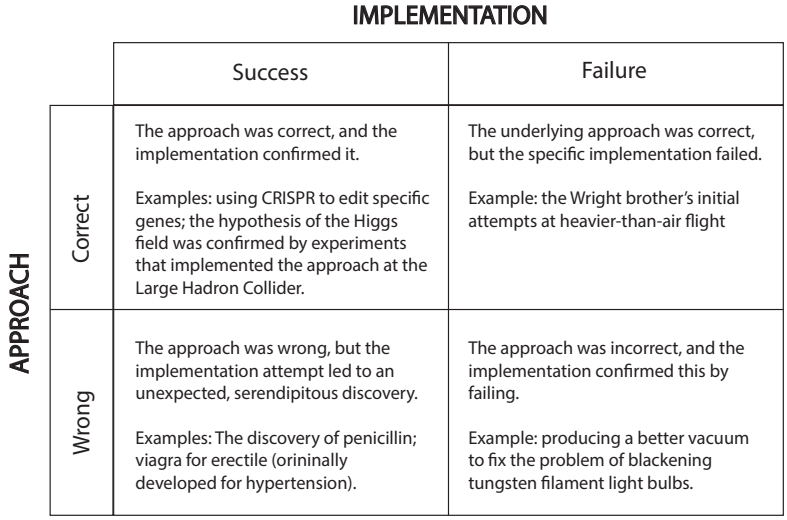
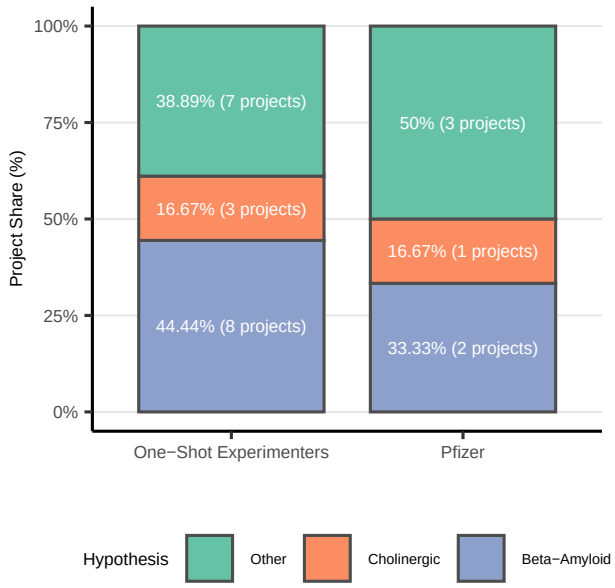


FIGURE 2: AD PROJECTS STARTED BY HYPOTHESIS FOR PFIZER AND SINGLE-SHOT FIRMS 2007-2008



Notes: This Figure compares the allocation of hypotheses across projects started by Pfizer to projects started by single-experimenter firms in 2007 and 2008. Single-experimenter firms are those that started only one AD drug development project in the period 2007-2008. We include projects where we can identify the year in which preclinical development started and the target used.

FIGURE 3: A DIRECTED ACYLIC GRAPH (DAG) CONNECTING THE THREE PROPOSITIONS

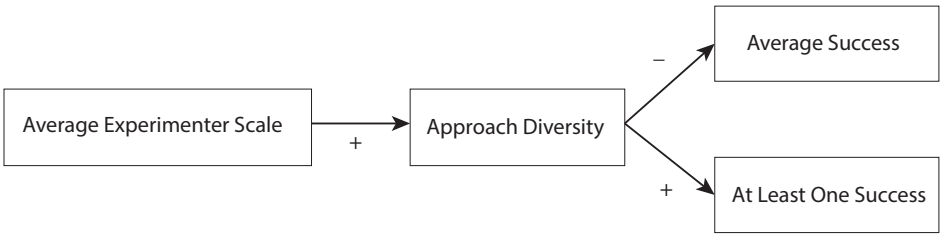
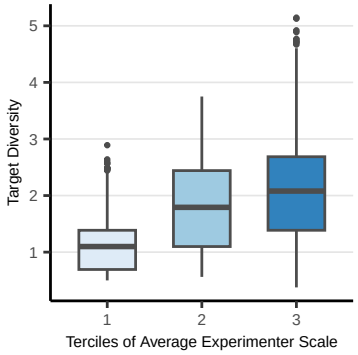
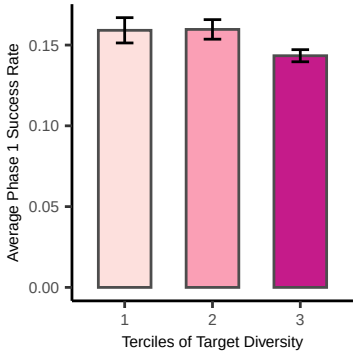


FIGURE 4: THE MARKET-LEVEL RELATIONSHIP BETWEEN EXPERIMENTER SCALE, THE DIVERSITY OF APPROACHES, AND SUCCESS

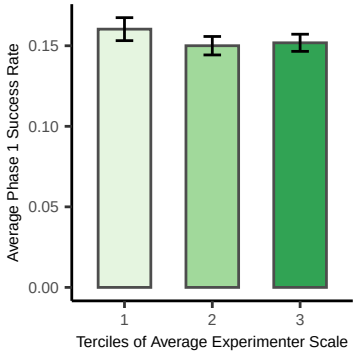
(A) Target Diversity & Average Experimenter Scale



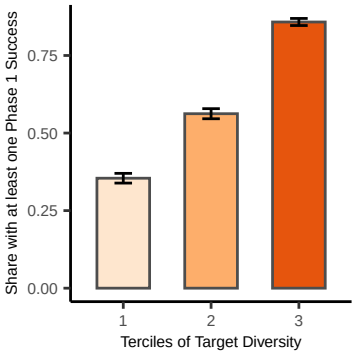
(B) Average Success & Target Diversity



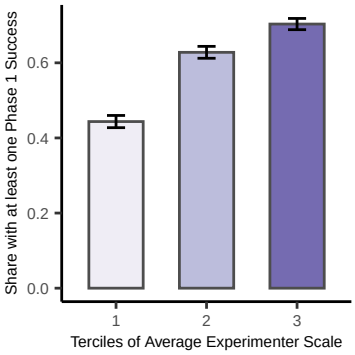
(C) Average Success & Average Experimenter Scale



(D) At Least One Success & Target Diversity



(E) At Least One Success & Average Experimenter Scale



Notes: This figure presents non-parametric plots corresponding to the parametric estimates in Tables 5 and 6. Panel (A) shows boxplots of target diversity by quintile of average experimenter scale. Panels (B) and (C) show bar charts of average Phase 1 success by quintiles of target diversity and experimenter scale, respectively. Panels (D) and (E) show the share of market-year observations with at least one Phase 1 success by quintiles of target diversity and experimenter scale, respectively. Error bars in panels (B)–(E) indicate standard errors.

FIGURE 5: A DIRECTED ACYLIC GRAPH (DAG) CONSIDERING THE EFFECT OF NEW TARGET-THERAPEUTIC CLASS DISCOVERIES

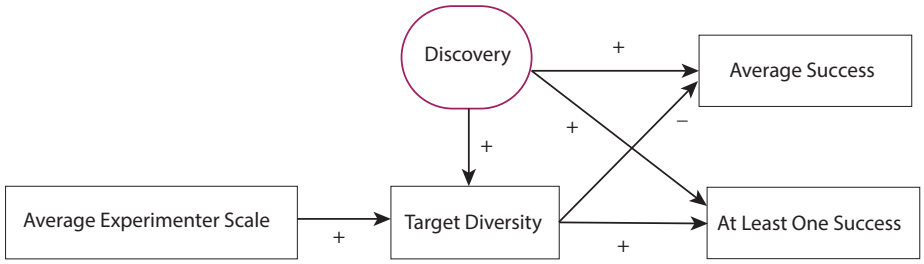
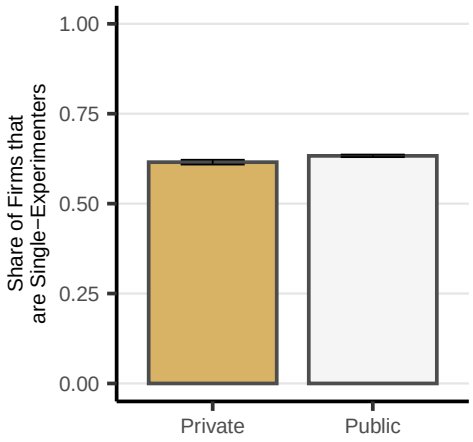
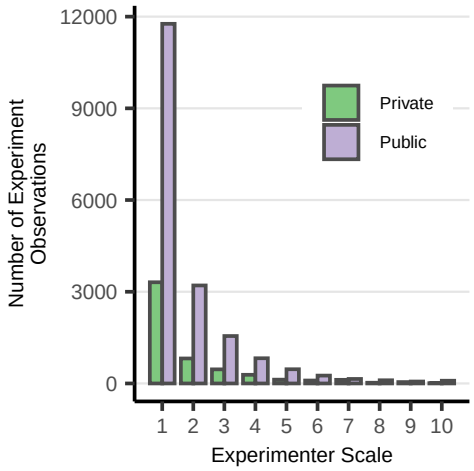


FIGURE 6: THE RELATIONSHIPS BETWEEN OWNERSHIP, EXPERIMENTER SCALE, AND THE LIKELIHOOD OF EXPERIMENTING WITH NOVEL TARGETS

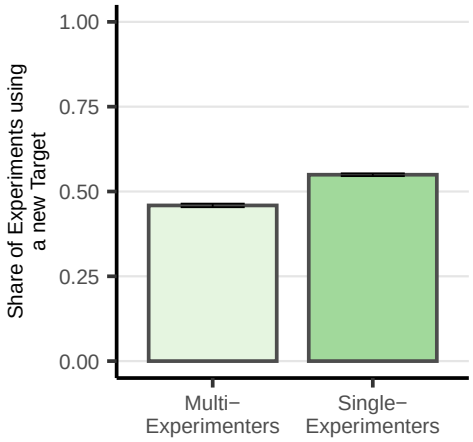
(A) Share of Single-Experimenter Firms by Ownership



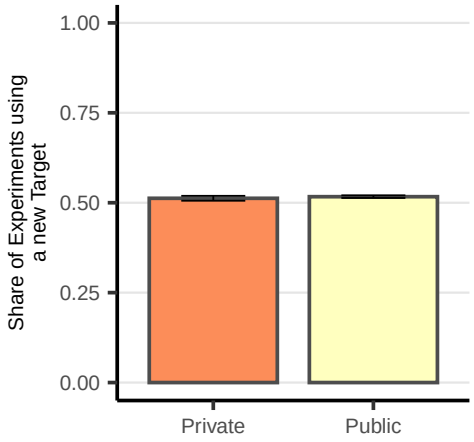
(B) Frequency of Experiments by Experimenter Scale



(C) Share of Novel Experiments by Experimenter Scale



(D) Share of Novel Experiments by Ownership



Notes: This figure examines non-parametric relationships between ownership, experimenter scale, and experimentation with novel targets, using data from 2005 onward to avoid inflated novelty due to the 1995 data start. Panel (A) compares the share of public and private firms conducting multiple experiments within a therapeutic class–year. Panel (B) shows experiment counts by experimenter scale for public and private firms. Panel (C) compares the share of novel-target experiments by multi- vs. single-experiment firms. Panel (D) compares the share of novel-target experiments by ownership type.

Table 1: Drug development projects started for Alzheimer's disease by hypothesis 1998-2008

Hypothesis	Number of Projects	Project Share (%)
Beta-Amyloid	49	39.2
Cholinergic	22	17.6
Other	54	43.2

Notes: This table shows the distribution of drug development projects in Alzheimer's Disease between 1998 and 2008 by hypothesis. We only include projects where we can identify the year in which preclinical development started and the target used. Our data describe the target that a drug is intended to act upon, not the broader underlying theory or hypothesis. For illustration purposes, in this table, we group targets into broader hypothesis groups. The beta-amyloid hypothesis includes drugs that target amyloid-beta, beta-secretase, adrenoceptor beta 1 and 2, and glutamate metabotropic. The cholinergic hypothesis includes drug projects that target cholinergic receptors and acetylcholine. Other hypotheses for AD include the tau hypothesis, mitochondrial cascade hypothesis, calcium homeostasis hypothesis, neurovascular hypothesis, inflammatory hypothesis, metal ion hypothesis, and lymphatic system hypothesis (Liu et al., 2019).

Table 2: Summary of Propositions

1	Holding the number of experiments constant, a firm pursuing multiple experiments will explore a greater variety of approaches compared to an equivalent number of projects, each pursued by a separate firm.
2	Markets with a higher share of small-scale experimenters will feature a lower diversity of approaches and will therefore have a higher share of successful experiments.
3	When higher experimenter scale is associated with greater diversity of approaches, markets with greater diversity of approaches have a higher probability of at least one success.

Table 3: Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Target Diversity	1.776	0.857	0.377	1.099	1.609	2.369	5.144	2523
Average Experimenter Scale	1.225	0.374	1.000	1.000	1.091	1.333	6.000	2523
Multi-Experiment Share	0.138	0.170	0.000	0.000	0.079	0.250	1.000	2523
ln(Discovery)	3.335	1.834	0.000	2.079	3.401	4.663	7.675	809
At least 1 Pre-Clinical Success	0.621	0.485	0.000	0.000	1.000	1.000	1.000	2523
Share of Pre-Clinical Success	0.155	0.182	0.000	0.000	0.111	0.250	1.000	2523
Project Start Year	2010.126	7.879	1996.000	2004.000	2010.000	2017.000	2023.000	2523
Number of Firms	8.897	12.276	2.000	3.000	5.000	10.000	165.000	2523
Number of Projects Started	11.568	19.329	2.000	3.000	6.000	13.000	306.000	2523
Number of Targets	9.782	14.772	2.000	3.000	5.000	11.000	210.000	2523

Notes: This table presents descriptive statistics for all variables in our therapeutic class–year (i, t) dataset featured in the main paper, of which there are $N = 2,523$ observations. Note that only 809 observations have values for *ln(Discovery)* due to the matching required between Pharmaprojects and GWAS (See Appendix D.4 for more details).

Table 4: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Target Diversity	1.00									
Average Experimenter Scale	0.28	1.00								
Multi-Experiment Share	0.26	0.82	1.00							
ln(Discovery)	0.24	0.02	-0.01	1.00						
At least 1 Pre-Clinical Success	0.46	0.19	0.17	0.12	1.00					
Share of Pre-Clinical Success	-0.02	0.06	0.06	-0.04	0.62	1.00				
Project Start Year	-0.02	-0.06	-0.10	0.46	-0.07	-0.04	1.00			
Number of Firms	0.79	0.29	0.26	0.22	0.31	0.02	0.08	1.00		
Number of Projects Started	0.73	0.37	0.31	0.19	0.27	0.04	0.07	0.98	1.00	
Number of Targets	0.77	0.34	0.29	0.20	0.28	0.03	0.07	0.99	0.99	1.00

Notes: This table details the correlation coefficient between all variables in our therapeutic class–year (i, t) dataset featured in the main paper, of which there are $N = 2,523$ observations. Note that only 809 observations have values for *ln(Discovery)* due to the matching required between Pharmaprojects and GWAS (See Appendix D.4 for more details).

Table 5: The market-level relationship between experimenter scale and the diversity of approaches

	Target Diversity		
	(1) OLS	(2) OLS	(3) OLS
Average Experimenter Scale	0.654***	0.682***	0.651***
	(0.173)	(0.188)	(0.093)
Market Structure Controls	No	No	Yes
ATC-1×Year FE	No	Yes	Yes
Observations	2,523	2,503	2,503
Adj R-squared	0.081	0.124	0.754

Notes: This Table reports regression results testing Proposition 1. In column (1) we regress *Target Diversity* on *Average Experimenter Scale*. In column (2), we add ATC-1×Year fixed effects, and in column (3), we include the set of *Market Structure Controls*. *Target Diversity* is the Shannon entropy of the relative abundance of targets employed in pre-clinical experiments, and *Average Experimenter Scale* is the average number of pre-clinical experiments started by firms in a therapeutic class–year. In all models, robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table 6: The market-level relationship between experimenter scale, the diversity of approaches, and success

	Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) 2SLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS
Target Diversity	-0.005	-0.045***		0.330***	0.312***	
	(0.011)	(0.012)		(0.024)	(0.036)	
Average Experimenter Scale			-0.029***			0.203***
			(0.005)			(0.044)
Market Structure Controls	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,503	2,503	2,503	2,503	2,503	2,503
Adj R-squared	0.060	-0.012	0.063	0.219	0.196	0.133
First Stage F-statistic		48.557			48.557	

Notes: This Table reports regression results testing Propositions 2 and 3. In columns (1) through (3), we test Proposition 2. In column (1) we report the OLS results from regressing the *Share of Pre-Clinical Success* on *Target Diversity*, where *Target Diversity* is the Shannon entropy of the relative abundance of targets employed in pre-clinical experiments. In column (2), we report the second stage results where *Target Diversity* has been instrumented with *Average Experimenter Scale*, where *Average Experimenter Scale* is the average number of pre-clinical experiments started by firms in a therapeutic class–year (the first stage results for this estimation are shown in Table 5 column (3)). Column (3) reports the OLS results from regressing *Share of Pre-Clinical Success* directly on *Average Experimenter Scale*. Lastly, columns (4) through (6) correspond to Proposition 3, where we explore the relationship between *Target Diversity* and *At least 1 Pre-Clinical Success*. Column (4) reports the OLS results, and column (5) reports second-stage estimates where, consistent with column (2), *Target Diversity* has been instrumented with *Average Experimenter Scale*. Column (6) regresses *At least 1 Pre-Clinical Success* on *Average Experimenter Scale*. In all models, robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table 7: Summary of Alternative Explanations

Alternative Explanation	**Argument Against**
If some markets have more technical opportunities, and this is positively (negatively) associated with the number of approaches observed, then there should be a positive (negative) relationship between diversity and average success but also at least one success. There is no reason to expect a relationship between diversity and average experimenter scale.	Results in Table 6 indicate a positive relationship between target diversity and at least one success, but a negative relationship between target diversity and the share of success
If experimenter-scale is positively related to technological opportunity or to superior firm capability, which, in turn, is positively related to the diversity of approaches, this would result in a positive association between experimenter-scale and diversity.	Experimenter scale would be positively associated with average success, contrary to the results in column (1) through (3) of Table 6.
The discovery of new targets and their relationships with therapeutic indications could expand the pool of available approaches, naturally increasing target diversity. Additionally, this dynamic could influence the composition of experimenter scale in the market if single- or multi-experiment firms differ in their likelihood of introducing novel targets.	Multi-experimenters are less likely to introduce new targets. If the discovery of new targets were the driver of diversity, then experimenter scale would be negatively associated with diversity, contrary to the results in Table 5. Further, diversity would then also be positively associated with average success, contrary to Table 6 columns (1) through (3). Finally, our results appear robust to controlling explicitly for rate of target discovery, as shown in column (1) through (4) of Table 8.
Younger firms may be less likely to conduct multiple experiments due to limited resources to fund and manage them. Additionally, if younger firms are more inclined to pursue novel approaches, the average firm age in a market could be correlated with target diversity.	Table 8 columns (5) through (8) show that our baseline results from Tables 5 and 6 are robust to the inclusion of our control for average firm age.

Table 8: The relationship between experimenter scale and target diversity while controlling for target discovery and firm age

	Target Diversity		Share of Pre-Clinical Success	At least 1 Pre-Clinical Success	Target Diversity		Share of Pre-Clinical Success	At least 1 Pre-Clinical Success
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
Average Experimenter Scale	0.736***	0.724***			0.680***	0.677***		
	(0.094)	(0.092)			(0.087)	(0.082)		
Target Diversity			-0.017	0.309***			-0.003	0.334***
			(0.014)	(0.020)			(0.010)	(0.023)
ln(Discovery)		0.032**	0.001	0.006				
		(0.013)	(0.006)	(0.015)				
ln(Average Firm Age)						0.066***	-0.005	-0.017
						(0.014)	(0.008)	(0.015)
Market Structure Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	2,422	2,422	2,422	2,422
Adj R-squared	0.768	0.770	-0.006	0.210	0.755	0.759	0.061	0.216

Notes:
This Table reports regression results testing each Proposition, where we also control for the discovery of new targets and firm age. In columns (1) and (2), we test Proposition 1 by regressing *Target Diversity* on *Average Experimenter Scale*, where the difference between the two models is that column (2) controls for discovery. *Target Diversity* is the Shannon entropy of the relative abundance of targets employed in pre-clinical experiments, and *Average Experimenter Scale* is the average number of pre-clinical experiments started by firms in a therapeutic class–year. *ln(Discovery)* is the natural logarithm of the count of publications in GWAS in year t that are the first to report a relationship between a target and therapeutic class i. In column (3), we test Proposition 2 and report the OLS results from regressing the *Share of Pre-Clinical Success* on *Target Diversity*. In column (4), we test Proposition 3, which concerns the relationship between *Target Diversity* and *At least 1 Pre-Clinical Success*. All models include the set of *Market Structure Controls*, and robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table 9: The market-level relationship between experimenter scale, the diversity of approaches, and success—with unproven targets only

	Target Diversity		Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) OLS
Average Experimenter Scale	0.879***	0.802***			-0.029***			0.251***
	(0.217)	(0.131)			(0.007)			(0.051)
Target Diversity			-0.004	-0.036**		0.325***	0.313***	
			(0.012)	(0.013)		(0.025)	(0.032)	
Market Structure Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,243	2,243	2,243	2,243	2,243	2,243	2,243	2,243
Adj R-squared	0.139	0.743	0.067	-0.009	0.069	0.217	0.197	0.131
First Stage F-statistic				37.334			37.334	

Notes: These table replicate the baseline results from Tables 5 and 6, where the underlying dataset has been modified to exclude projects using targets which have been targeted by at least one launched drug in the same therapeutic class. Robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Appendix:

If you had only one shot: Scale and herding in innovation experiments

A A Case Study on Approaches to Experimentation: Psoriasis

Psoriasis[26] is a chronic inflammatory skin condition affecting approximately 2-3% of the global population (Ayala-Fontanez et al., 2016). A common symptom of the disease is the appearance of raised, silvery plaques (Nestle et al., 2009). Although treatments for psoriasis have been available for decades, the disease has many forms, and the causes of some less common types are still not fully understood (Guo et al., 2023). Many hypotheses exist, and some examples include: (i) overactive T-cells triggering inflammation and rapid skin cell production, (ii) genetic factors, such as mutations in the HLA-Cw6 gene, (iii) involvement of cytokines like IL-23 and IL-17, (iv) inflammatory lipid molecules called leukotrienes, and (v) abnormalities in keratinocytes that result in excessive skin cell production. This section describes examples of firms initiating pre-clinical trials for two distinct psoriasis drug development projects within the same year. These examples are sourced directly from Pharmaprojects and supplemented with details from Trialtrove.

Astrazeneca. In 2004, AstraZeneca began preclinical trials for two psoriasis treatments. One was a humanized antibody called Sifalimumab, which targeted interferon-alpha (IFN-α). The underlying hypothesis was that in genetically predisposed individuals, the immune system is primed, and exogenous IFN-α may trigger psoriasis development.

At the same time, pre-clinical trials began for Certolizumab pegol, a recombinant humanized high-affinity anti-TNFalpha antibody fragment, developed by UCB (Celltech before the acquisition), for the treatment of chronic inflammatory conditions, including Crohn's disease (CD), rheumatoid arthritis (RA), psoriatic arthritis and ankylosing spondylitis

The anti-TNF alpha psoriasis hypothesis suggests blocking Tumor Necrosis Factor-alpha (TNF-alpha). It has since been shown, however, that this approach can lead to the development or worsening of psoriasis, primarily due to an uncontrolled increase in type 1 interferons

[26] https://www.niams.nih.gov/health-topics/psoriasis

produced by plasmacytoid dendritic cells (pDCs), which are key players in psoriasis pathogenesis.

Stiefel Laboratories (GlaxoSmithKline). In 2006, Stiefel Laboratories began preclinical trials for two psoriasis drugs. One was called Primolux, which was a 0.05% topical formulation of the corticosteroid clobetasol, developed using its proprietary VersaFoam-EF technology, that targeted the nuclear receptor subfamily 3 group C member 1. The second drug, Calcipotriol VersaFoam, was a vitamin D receptor antagonist. The treatment consisted of a 0.005% topical formulation of calcipotriol, a vitamin D3 analog.

In 2009, GlaxoSmithKline acquired Stiefel Laboratories for $2.9 billion to create a specialist dermatology branch. [27] Consequently, in Pharmaprojects, these projects are described as having originated by GlaxoSmithKline. This example highlights the complications in measurement. Pharmaprojects assigns the "company developing a drug" to each drug-treatment. Importantly, this would suggest that the listed focal company is both funding development and the primary decision maker.

Identifying the impact of such measurement error is hard to do. In the example of Stieffel Laboratories, this measurement error would not affect econometric estimates in our core results, at least. We would still consider the experiments to be done by a multi-experiment firm. If, however, we are on average more likely to incorrectly assign multi-experiment status to a large incumbent such as GlaxoSmithKline, then our analysis in Section 6.3 may inaccurately estimate the relationship between ownership type, the scale of experimentation, and the introduction of novelty (Figure 6 panels (A), (B), and (D))

However, measurement error in general would create an attenuation bias. In particular, if measurement error is also correlated with target diversity. For example, if firm X acquired two single experimenters, each with a distinct approach, and both experiments are attributed to firm X, then this would be a positive correlated measurement error leading to an upward bias on the estimated coefficient. Given our data, it is not feasible to scrutinize the history

[27] https://www.gsk.com/en-gb/media/press-releases/glaxosmithkline-completes-acquisition-of-stiefel

of each drug development project and the accuracy of the originator firm assignment. This should be recognised as an empirical limitation of this study.

B Model Extensions

While simulations are best suited to generalizing our model (see Appendix C), we can formally extend the model to analyze the decision of the n^{th} experiment after $n-1$ experiments in approach a. For simplicity, we set $\pi_a = \pi_b = \pi$. Instead, approach a is assumed to be more promising than b because $p_a > p_b$ for all experimenters.

In addition, we relax our rent-sharing assumption. In the baseline model in the two-firm case, we assume that if both experiments are successful, only one experiment will capture value (with probability 0.50). In this extension, we consider the scenario where the probabilities of value capture can be lower. Concretely, this permits the scenario where two firms are successful, and they can both commercialize their innovations, but due to competition, the value they capture is less than $V/2$. We model rent dissipation with the parameter k to analyze the effect of rent dissipation on the choice of approach.

Suppose there are n firms and n experiments in approach a and 0 experiments in b, such that each firm is a single-experiment firm. A potential entrant in approach a has a payoff of:

$$\begin{aligned} & p_a\pi\left((1-p_a)^n + \frac{1}{2+k}\binom{n}{1}a(1-p_a)^{n-1} + \frac{1}{3+k}\binom{n}{2}(p_a)^2(1-p_a)^{n-2}...\frac{1}{n+1+k}(p_a)^n\right) - c \\ & = p_a\pi X(n,k) - c \end{aligned} \tag{4}$$

Note that $X(n,k)$ is strictly less than unity, and decreases with the rent dissipation parameter k, because each of the terms $\frac{1}{n+1_k}p_a^n$ decreases with k. Note also that $X(n,k) - (1-p_a)X(n-1,k)$ is positive but falls with k. Formally,

$$\begin{aligned} & X(n,k) - (1-p_a)X(n-1,k) = \\ & \frac{1}{2+k}\left(\binom{n}{1} - \binom{n-1}{1}\right)a(1-p_a)^{n-1} + \frac{1}{3+k}\left(\binom{n}{2} - \binom{n-1}{2}\right)p_a^2(1-p_a)^{n-2} \\ & .. + \frac{1}{n+k}(p_a)^n > 0 \end{aligned} \tag{5}$$

That $X(n,k) - (1-p_a)X(n-1,k)$ decreases with k follows upon noting that each term $\frac{1}{r+k}\left(\binom{n}{r-1} - \binom{n-1}{r-1}\right) p_a^{r-1}(1-p_a)^{n-1-(r-1)}$, decreases with k.
Entering b instead has a payoff of:

$$p_b\pi\left(\pi X + (1-\pi)\right) = p_b\pi^2 X(n,k) + p_b\pi(1-\pi) - c \tag{6}$$

The incentive to herd into approach a for the single-experiment entrant is given by:

$$\Delta_s = \pi(p_a - \pi p_b)X(n,k) + \pi(1-\pi p_b) \tag{7}$$

Now consider a potential multi-experiment firm, which has one experiment in a along with $n-1$ other firms i.e., an incumbent. The payoff to this firm of a second experiment in a is:

$$\begin{aligned} &\pi(p_a(2-p_a)\left((1-p_a)^{n-1} + \frac{1}{2+k}\binom{n-1}{1}p_a(1-p_a)^{n-2} + ...\frac{1}{n+k}p_a^{n-1}\right) \\ &= \pi X(n-1,k)p_a(2-p_a) - c \end{aligned} \tag{8}$$

And their payoff from entering with approach b is given by:

$$\pi X(n-1,k)(a + \pi(1-a)b) + b\pi(1-\pi) - c \tag{9}$$

Thus the incentive to herd in approach a for the incumbent firm can be expressed:

$$\Delta_M = \pi(1-p_a)X(n-1,k)(p_a - \pi p_b) + \pi(1-\pi)p_b \tag{10}$$

It follows that $\Delta = \Delta_s - \Delta_M = \pi(1-\pi p_b)\left[X(n,k) - (1-p_a)X(n-1,k)\right] > 0$. That is, if the $(n+1)^{th}$ experiment is conducted by a new entrant, they will be more likely to experiment with approach a compared to a multi-experiment firm who is incumbent with an existing experiment in approach a. Furthermore, Δ falls with k because $X(n,k) - (1-p_a)X(n-1,k)$ falls with k. **That is, small-scale experimenters are more likely to herd than large-**

scale experimenters, but less so when rents are dissipated.

C Numerical Simulations

C.1 Baseline

Set up. In the model developed in Section 4, there are two important assumptions underlying its basic intuition. First, the total value that can be captured by successful firms is independent of the number of successful firms, meaning the total profit from successful experiments by independent firms stays constant. Specifically, a firm with two successful experiments earns the same payoff as if it had only one successful experiment. This implies that competition among successful firms does not dissipate rents but merely distributes the total payoff in some fashion. The second assumption is that the validity of approaches is uncertain. Firms choosing the same approach are more likely to succeed or fail together than firms choosing different approaches, meaning value diversion is more likely when firms choose the same approach. In other words, firms choosing the same approach have correlated outcomes, whose effect is not fully integrated into individual firms' decisions. However, this correlation arises only if $\pi \neq 1$; if $\pi = 1$, the probability of success depends solely on implementation. Thus, firms will choose the approach with which they have the highest probability of implementation success. Target diversity at the market level would simply reflect differences in implementation ability for different approaches.

Generalizing beyond two approaches and two possible experiments to multiple approaches and multiple firms involves considerations of competition and order of entry into experimentation. While this is not analytically tractable, simulations permit us to explore the robustness of the model's intuition. We use Monte Carlo simulations to generalize the results of our model. In the baseline simulation, we define a pool of 50 firms, of which $\phi \in [0, 1]$ have a high cost of experimenting (and $1 - \phi$ incur a low cost). In each period, a firm is randomly selected and can choose to experiment. Firms experiment with the approach that returns the largest non-negative expected payoff and do not enter if all expected payoffs are negative. This payoff depends on the ongoing experiments and whether the firm has an

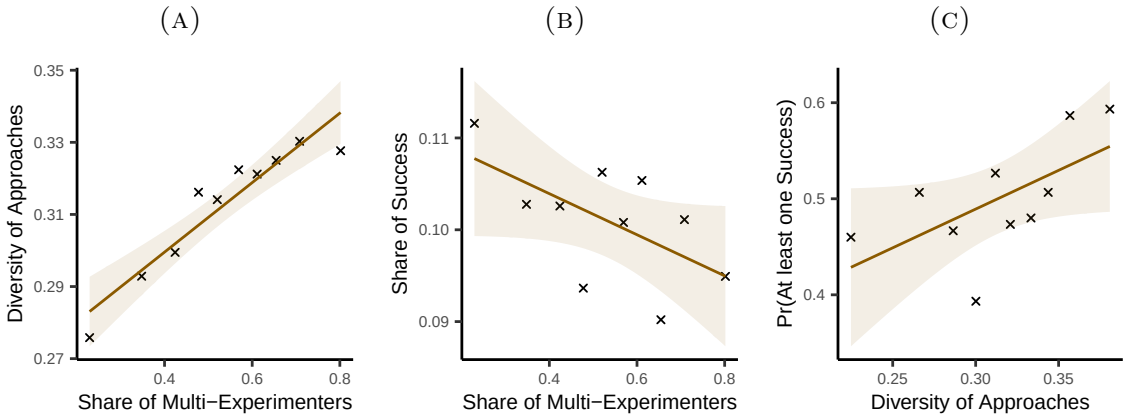
ongoing experiment. Once a firm has entered twice, we remove it from the pool of firms. We run simulations for different shares ϕ to exogenously create variation in the share of firms that enter with two experiments (a higher share of high-cost firms leads to fewer firms entering with two experiments). A simulation ends when the first low-cost firm chooses not to enter, since this indicates that there will be no more entry (high-cost firms are less likely to enter than low-cost firms, given their higher cost). We run 400 simulations for $\phi = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$

Results. Figure C.1 presents results from the baseline simulation, where we simulate firms choosing between two approaches, a and b, where $\pi_a < \pi_b$. Firms continue to enter until they become unprofitable with either approach. In panel C.1a, we link the market-level share of multi-experiment firms to the diversity of approaches used. Maximum diversity would be an even split between approach a and b, with target diversity defined as one minus the share of experiments using approach a. Our simulations support the analytical prediction that greater approach diversity occurs when firms conduct multiple experiments on average. In panels C.1b and C.1c, we show results from simulations predicted by Propositions 2 and 3. Markets with a greater share of multi-experiment firms have a lower individual experiment success rate. However, in markets with greater approach diversity—associated with more multi-experiment firms—we observe a higher probability of at least one experiment succeeding.

C.2 Discovery of new Approaches

A feature of our empirical context is that new approaches are discovered. Additionally, in practice, more than two approaches will be available to the experimenter. In our empirical context, for example, the average number of targets observed in a therapeutic class–year is 9.09 (median=5, std.dev.=14.21, min=2, max=210. Here, we extend our simulation framework to allow for discovery and, thus, multiple approaches.

FIGURE C.1: SIMULATION RESULTS

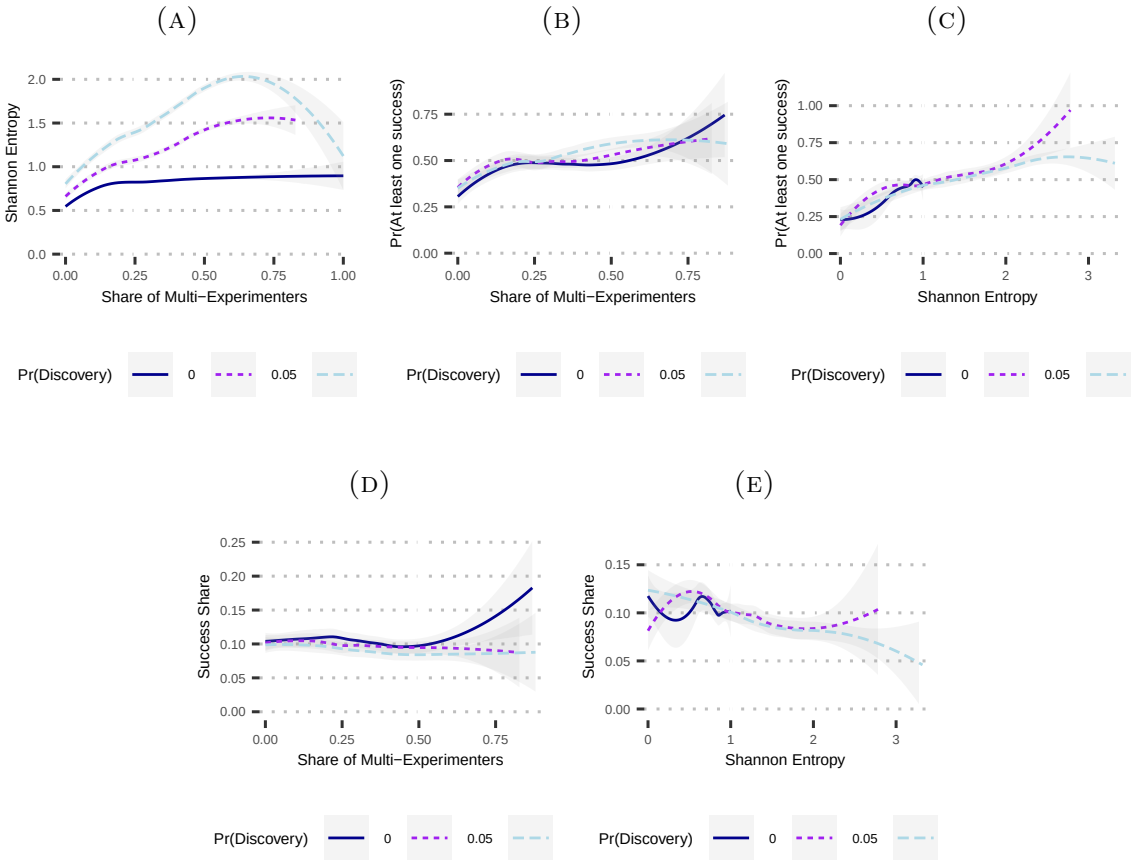


Notes: These figures report binned scatter plots of our simulation results, illustrating the relationships predicted by each proposition. We simulate 1500 markets, creating variation in the share of large-scale experimenters by exogenously shifting the share of firms that have a high cost of experimenting. A detailed explanation of the simulation logic is provided in Appendix C. This simulation uses the following parameters: $\pi_a = 0.4$, $\pi_b = 0.2$, $p_a = p_b = 0.3$.

Set up. The logic for these simulations is identical to that described in the baseline simulations, but we make the following changes to introduce discovery. In each period, a new approach is discovered with probability γ. In the case of discovery, a new entrant enters with the approach, and the approach enters the domain of available approaches that subsequent experiments can choose from. We keep all parameter values the same as the baseline simulation (i.e., $\pi_a = 0.4; \pi_b = 0.2; p_a = p_b = 0.3$). All new approaches k have a lower viability probability $\pi_k = 0.1$, but the same implementation probability $p_k = 0.3$ as existing approaches a and b. We assume that only new entrants discover new targets and that this is their only experiment. This is to match our empirical finding, that the introduction of new approaches is more likely to happen with single-experiment firms. Keeping all other parameters constant, we vary the probability of discovery and observe the effect of discovery on the diversity of targets and success. We run 400 simulations for each combination of high-cost experimenter share $\phi = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ and discovery probability $\gamma = \{0, 0.05, 0.1\}$.

Results. In Figure C.2, we report simulation results grouped by each discovery probability. We measure the diversity of approaches with Shannon entropy, consistent with our empirical

FIGURE C.2: Simulating the Effect of Discovery on the Diversity of Approaches and the Outcomes of Experiments



method. In panel (A), our simulations show that the probability of discovery positively moderates the relationship between the share of multi-experimenters and the diversity of approaches. In other words, for a fixed share of multi-experiment firms, when the likelihood of a new approach being discovered is higher, our simulations predict a greater diversity of approaches.

Turning to outcomes, the results are less clear-cut. In panels (B) and (C), we look at the probability of at least one success, and in panels (D) and (E), we look at the average success rate. In panel (B), the simulations suggest a small benefit from discovery on the probability of at least one success. This is consistent with our previous results: a higher discovery rate is associated with a greater diversity of approaches, which is associated with a

higher probability of the market finding at least one success. Panel (C) is harder to interpret, but our results suggest a declining benefit to higher discovery rates. Holding the diversity of approaches constant, a higher probability of at least one success is achieved when the discovery rate is lower.

Panel (D) also conforms with prior results. The share of successful experiments is lower when a market consists of a greater share of multi-experiment firms, especially when the discovery rate is higher. Discovery leads to new approaches becoming available, but individually, they have a smaller probability of success. Panel (E) suggests results consistent with the intuition of the panel (C). Holding the diversity of approaches constant, the share of experimental successes will be lower when the discovery rate is higher.

C.3 Market Equilibrium

Overview. The empirical exercise requires recovering an unobserved, latent share of multi-experiment firms for each therapeutic class–year. Numerical imulations allow us to define a latent population and characterize the equilibrium distribution of experiments that would arise under our model. This provides two benefits: (i) it clarifies the steady-state predictions of the model once we move beyond the two-firm, two-approach analytical setting, and (ii) it enables calibration of the model's key parameters to empirical moments.

Set up. We consider two approaches, a and b, with approach-level viability probabilities $\pi_a = 0.4$ and $\pi_b = 0.2$, and conditional implementation success probabilities $p_a = p_b = 0.5$. The market contains 20 firms, of which 9 have experimentation capacity 2 ('multi-experimenters') and the remaining 11 have capacity 1 ('singletons'). Capacity determines the maximum number of experiments a firm can perform. Firms observe the complete history of prior entries, (S_a, S_b), before choosing whether and where to enter, but they do not update beliefs; approach viability remains fixed at (π_a, π_b).

In each period a firm with remaining capacity is drawn uniformly at random and decides

whether to experiment. A capacity–1 firm chooses the approach $t \in \{a, b\}$ with the highest expected payoff,

$$\mathbb{E}[\Pi_t p_t \cdot \tfrac{1}{1+K_{-i}}]\, V - c,$$

where K_{-i} is the random number of rival successes, integrated over both approach viability and binomial implementation uncertainty. A capacity–2 firm makes a second decision after its first experiment: it chooses either to repeat the same approach or to diversify into the other approach. The expected payoff from the second experiment accounts for (i) the probability that at least one of the firm's two attempts succeeds, and (ii) the expected rent share conditional on rival successes, holding fixed the firm's first experiment.

Entry occurs only if the relevant expected payoff is positive; otherwise the firm abstains. Firms exit the active pool when capacity is exhausted. Rents are split equally among all successful firms. Simulations continue until either all firms have exhausted capacity or ten consecutive firms decline to enter, which serves as a practical equilibrium stopping rule. Finally, we vary the value parameter $V \in \{1, 2, \ldots, 6\}$ and compute equilibrium statistics (approach shares, experiment success rates, and the participation rate of multi-experimenters), which are used for calibration against empirical moments (see Table C.1).

Table C.1: Parameters are calibrated to match empirical moments

	Simulations	Data
Average success (%)	17.4	15.5
Multi share (%)	13.9	13.8
N firms	10.2	8.90
N experiments	12.0	11.57

Figure C.3 reports the key equilibrium results. Panel C.3a verifies that in steady state, the share of experiments using approach a converges to the theoretical benchmark $\pi_a/(\pi_a+\pi_b) = 0.67$. This reflects the underlying viability parameters, rather than specific features of entry order or capacity constraints.

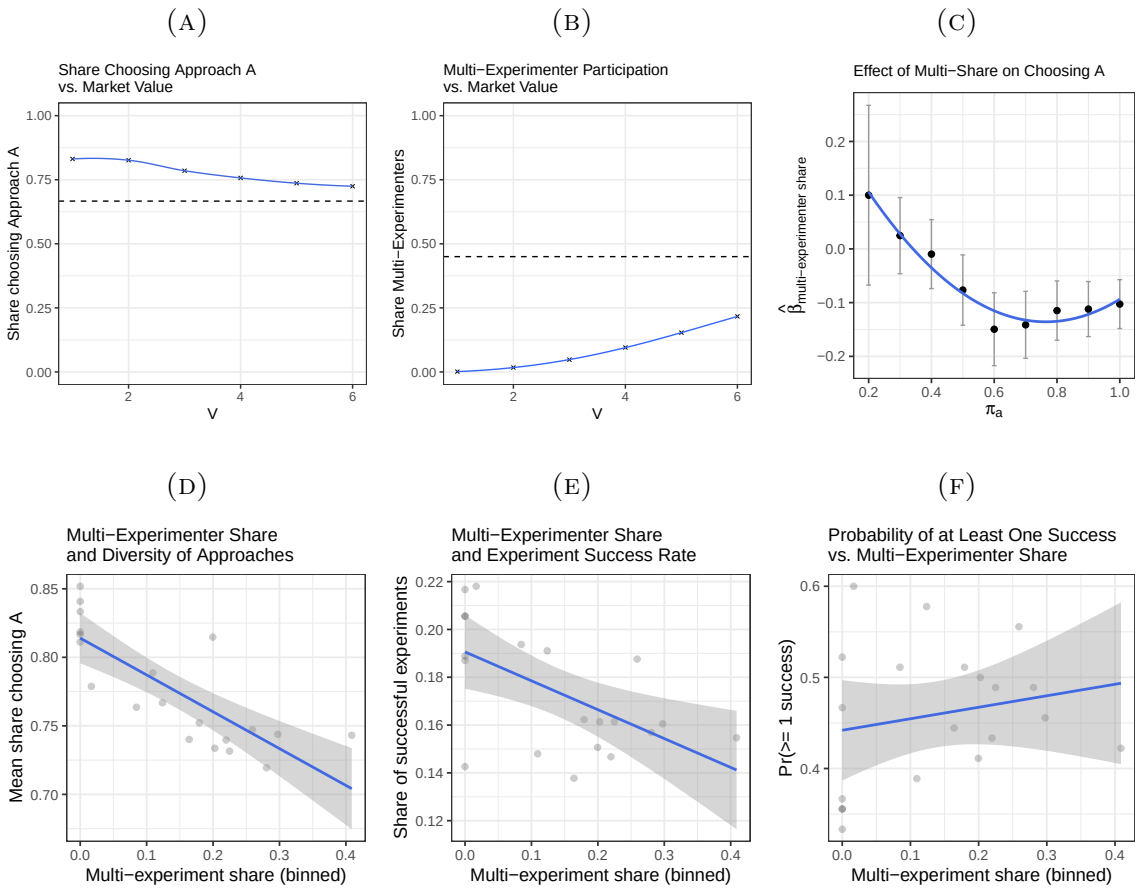
Panel C.3b shows that the observed share of multi-experimenters in equilibrium converges to the true population share as markets grow larger. This confirms that the simulation

recovers the correct latent structure.

Panel C.3c illustrates that the effect of multi-experimenter share on approach diversity depends on approach viability (π_a, π_b). The relationship is strongest at intermediate values of π, consistent with the model's intuition: diversity is most sensitive to experimenter scale when the correlation risk within an approach is moderate.

Lastly, panels C.3d, C.3e, and C.3f mirror the comparative statistics derived from the model. Multi-experiment share is negatively associated with share of experiments in A (positive with diversity), negative with average success, and positive with at-least-one success.

FIGURE C.3: Simulating Market Equilibrium



Notes: In panel (C), the data are generated by running the simulation (same calibrated parameters) over different values of $\pi_a \in \{0.2, 0.3, \ldots, 1\}$, where $\pi_a = \pi_b - 0.2$. Marginal effects are estimated separately at each value of π_a from regressions of Approach A Share on Multi-Experimenter Share, controlling for the number of entries and V. Each point is the OLS coefficient on multi share at that π_a, with 95% confidence intervals. The blue curve is a smoothed fit across π_a. All simulations hold $\pi_a - \pi_b$ constant. In panels (D)–(F), each figure displays a binned scatter plot of the raw simulation outcomes. Data is binned by twenty equal chunks based on the distribution of multi-experiment share. A linear fit is overlaid on the binned means.

D Data

D.1 Description of Variables

D.2 Matching Pharmaprojects to Pitchbook

We match data on firm founding year and ownership from Pitchbook to Pharmaprojects by firm name. Matching by firm name is challenging because datasets often use different naming conventions. For example, one dataset may list "Eli Lilly & Co.", while the other simply states "Lilly". Here we briefly describe our approach to this inherently noisy matching task and report results from manual validation.

We first filter firms in Pitchbook by their primary industry group, selecting industries that contain the phrases pharmaceuticals, healthcare, drugs, medical, surgical, hospitals, and clinics. We then clean firm names, removing punctuation, special characters, and common legal suffixes, e.g., ltd., inc., co.

Matching on these cleaned firm names produces multiple matches for each firm in Pharmaprojects. For example, we find that Pfizer matches to many subsidiaries or overseas business units. To identify the focal firm in each match, we keep the matched firm with the oldest founding year. If this data is missing, we select the matched firm with the largest number of employees.

Out of 2,804 pharmaceutical firms, we matched 2,150 to Pitchbook. To validate our match, we took a random sample of 50 firms from those that were matched and another random sample of 50 firms from those that were not matched. For those that were matched, we compared the matches across other observable dimensions, such as HQ location. These manual checks revealed one error in the random sample of 50 firms. As for the unmatched firms, we repeated the same process but with the unfiltered Pitchbook data (i.e., not conditioning on pharmaceutical-relevant firms). For the 654 unmatched firms, we found a match for 76 firms. However, we found that only half of these were accurate upon manual inspection.

Variable Name	Description
Target Diversity	Shannon entropy of the relative abundance of targets employed in pre-clinical experiments.
HHI	Herfindahl-Hirschman Index of targets employed in pre-clinical experiments.
Average Experimenter Scale	Average number of pre-clincial experiments started by firms.
Average Experimenter Scale $(t-1)$	Average number of pre-clincial experiments started by firms in the previous year.
Multi-Experiment Share	Share of pre-clincial experiments started by firms who themselves launched two or more experiments.
Multi-Experiment Share $(t-1)$	Share of pre-clincial experiments started by firms in the previous year who themselves launched two or more experiments.
ln(Discovery)	Natural logarithm of the count of publications in GWAS that report a novel target.
Average Firm Age	Average age of distinct experimenters (firms).
At least 1 Pre-Clinical Success	A dummy variable equal to 1 if at least one pre-clinical trial is successful and is followed by a Phase 1 clinical trial; 0 otherwise.
Share of Pre-Clinical Success	Share of pre-clinical trials that are followed by a Phase 1 clinical trial.
At least 1 Drug Launch	A dummy variable equal to 1 if at least one pre-clinical trial is successful and ultimately results in the launch of a drug product, 0 otherwise.
Share of Drug Launch	Share of pre-clinical trials that eventually result in the launch of a drug product.
Project Start Year	Year in which experimentation begins. We define an experiment as the first pre-clinical trial in a drug development project.
Number of Firms	Total number of firms experimenting.
Number of Projects Started	Total number of pre-clinical experiments initiated.
Number of Targets	Total number of targets being drugged in experiments.

Notes: All variables are defined at the therapeutic class–year level (i, t). All variables are created using the Pharmaprojects dataset, except for *ln(Discovery) which is constructed using data from GWAS, and* Average Firm Age *which uses data from Pitchbook.*

In sum, our matching approach is highly accurate, but the main limitation is the coverage of the data.

Table D.1: Descriptive Statistics

	Mean	St. Dev.	Min	Q1	Median	Q3	Max	N
Target Diversity	1.776	0.857	0.377	1.099	1.609	2.369	5.144	2523
HHI	0.237	0.163	0.007	0.098	0.200	0.333	0.781	2523
Average Experimenter Scale	1.225	0.374	1.000	1.000	1.091	1.333	6.000	2523
Average Experimenter $\text{Scale}_{\text{t-1)}}$	1.225	0.370	1.000	1.000	1.100	1.333	6.000	2384
Multi-Experiment Share	0.138	0.170	0.000	0.000	0.079	0.250	1.000	2523
Multi-Experiment $\text{Share}_{\text{t-1)}}$	0.138	0.168	0.000	0.000	0.083	0.250	1.000	2384
ln(Discovery)	3.335	1.834	0.000	2.079	3.401	4.663	7.675	809
Average Firm Age	37.286	30.866	0.000	13.500	28.545	52.778	162.000	2447
At least 1 Pre-Clinical Success	0.621	0.485	0.000	0.000	1.000	1.000	1.000	2523
Share of Pre-Clinical Success	0.155	0.182	0.000	0.000	0.111	0.250	1.000	2523
At least 1 Drug Launch	0.170	0.376	0.000	0.000	0.000	0.000	1.000	2523
Share of Drug Launch	0.028	0.086	0.000	0.000	0.000	0.000	1.000	2523
Project Start Year	2010.126	7.879	1996.000	2004.000	2010.000	2017.000	2023.000	2523
Number of Firms	8.897	12.276	2.000	3.000	5.000	10.000	165.000	2523
Number of Projects Started	11.568	19.329	2.000	3.000	6.000	13.000	306.000	2523
Number of Targets	9.782	14.772	2.000	3.000	5.000	11.000	210.000	2523

Notes: This table presents descriptive statistics for all variables in our therapeutic class–year (i, t) dataset used in the main paper and robustness analysis featured in the Appendix.

D.3 Identifying ATC-1 groups

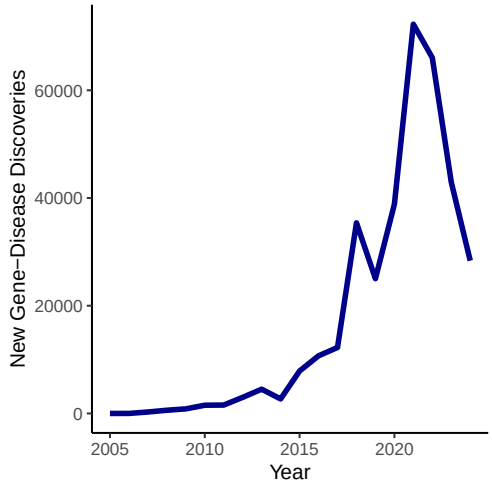
In our analyses, we control for unobserved time and industry varying trends using $ATC1 \times Year$ fixed effects (Branstetter et al., 2022). Our data, however, do not include the ATC-1 code to which the therapeutic class belongs. Fortunately, there are only 14 ATC-1 categories (`https://www.who.int/tools/atc-ddd-toolkit/atc-classification`). We exclude ATC P: Antiparasitic products, insecticides and repellents, which leaves 13 codes. In a similar approach to matching our sample to GWAS, we loop through our 200 therapeutic class names and use GPT-4.0 to identify the closest ATC-1 match. We match 139 out of 200 therapeutic class designations to an ATC-1 code. Importantly, the therapeutic classes that we match are the most pervasive, as we only lose 268 (out of 2,791) observations when we exclude therapeutic classes that we are unable to match to an ATC-1. We also ran all of our analyses with these 268 observations grouped under a catch-all ATC-1 'X'. Though not reported here, all of these results were very similar to those reported in the main paper and this Appendix.

D.4 Matching Pharmaprojects to GWAS

A key feature of our analysis includes controlling for the frequency of discovery. The variable we construct—ln(Discovery)—is the natural logarithm of the number of GWAS publications reporting a new target relevant to therapeutic class i in year t.

Founded in 2008 by the National Human Genome Research Institute, GWAS (`https://www.ebi.ac.uk/gwas`)—or the Catalog of human Genome-Wide Association Studies—is a record of all scientific publications in top-tier journals reporting the discovery of a new target–disease correspondence. An in-depth overview of GWAS is provided by Tranchero (2023).

Figure D.1: New Gene-Disease Discoveries Over Time



The GWAS catalog only reports the disease and/or trait that a particular study is concerned with, so to match these data to Pharmaprojects, we need to make a correspondence between disease/traits and therapeutic class. We first created two lists of the 200 distinct therapeutic class names and 28,540 disease names from GWAS. Next, we wrote a Python script that looped through each disease name and used OpenAI's GPT-4.0 API to identify any and all (at most three) corresponding therapeutic class names. We matched 22,381/28,840 disease names to 179/200 therapeutic classes. Then, for each therapeutic class–year, we count the number of new discoveries in corresponding diseases published in

GWAS in that year. As Figure D.1 shows, the coverage of GWAS is considerably weaker for the first three-quarters of our sample period (GWAS was only founded in 2008). Consequently, we have a measure of discovery for 811 out of 2791 market observations. After dropping observations where we don't have an ATC-1 code, this becomes 809.

Table D.2: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Target Diversity	1.00															
HHI	-0.89	1.00														
Average Experimenter Scale	0.27	-0.15	1.00													
Average Experimenter $\text{Scale}_{\text{t-1)}}$	0.22	-0.11	0.26	1.00												
Multi-Experiment Share	0.26	-0.15	0.82	0.22	1.00											
Multi-Experiment $\text{Share}_{\text{t-1)}}$	0.17	-0.06	0.21	0.81	0.20	1.00										
ln(Discovery)	0.24	-0.18	0.02	-0.02	-0.01	-0.02	1.00									
Average Firm Age	-0.05	-0.00	0.03	0.02	0.05	0.03	-0.15	1.00								
At least 1 Pre-Clinical Success	0.46	-0.43	0.19	0.13	0.16	0.11	0.12	-0.04	1.00							
Share of Pre-Clinical Success	-0.03	0.07	0.06	0.06	0.06	0.08	-0.04	-0.08	0.62	1.00						
At least 1 Drug Launch	0.22	-0.16	0.11	0.15	0.12	0.08	-0.05	0.07	0.15	0.09	1.00					
Share of Drug Launch	-0.10	0.10	-0.00	0.01	0.04	-0.03	-0.08	0.07	0.03	0.12	0.67	1.00				
Project Start Year	-0.02	0.08	-0.06	-0.04	-0.09	-0.05	0.46	-0.35	-0.07	-0.04	-0.26	-0.15	1.00			
Number of Firms	0.79	-0.49	0.30	0.29	0.26	0.24	0.22	-0.10	0.30	0.02	0.18	-0.06	0.08	1.00		
Number of Projects Started	0.73	-0.43	0.37	0.32	0.31	0.26	0.19	-0.09	0.27	0.04	0.17	-0.05	0.07	0.98	1.00	
Number of Targets	0.77	-0.47	0.34	0.31	0.29	0.25	0.20	-0.09	0.28	0.03	0.18	-0.06	0.07	0.99	0.99	1.00

Notes: This table presents descriptive statistics for all variables in our therapeutic class–year (i, t) dataset used in the main paper and robustness analysis featured in the Appendix.

E Robustness Analysis

Table E.1: The market-level relationship between experimenter scale, the diversity of approaches, and success—Alternate time windows

(a) Unit of Observation = Therapeutic Class—Two Year Period

	Target Diversity		Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) OLS
Average Experimenter Scale	0.730***	0.569***			-0.030***			0.138**
	(0.205)	(0.096)			(0.007)			(0.046)
Target Diversity			-0.003	-0.053***		0.288***	0.243***	
			(0.011)	(0.017)		(0.020)	(0.053)	
Market Structure Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,557	1,557	1,557	1,557	1,557	1,557	1,557	1,557
Adj R-squared	0.148	0.741	0.064	-0.032	0.069	0.259	0.211	0.157
First Stage F-statistic				35.234			35.234	

(b) Unit of Observation = Therapeutic Class—Five Year Period

	Target Diversity		Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) OLS
Average Experimenter Scale	0.954***	0.484***			0.008			0.127***
	(0.162)	(0.051)			(0.013)			(0.031)
Target Diversity			-0.008	0.016		0.157***	0.263***	
			(0.008)	(0.027)		(0.024)	(0.070)	
Market Structure Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	398	398	398	398	398	398	398	398
Adj R-squared	0.233	0.760	0.059	-0.025	0.058	0.132	0.082	0.070
First Stage F-statistic				89.413			89.413	

Notes: These table replicate the results from Tables 5 and 6, where the underlying dataset has been modified to allow for longer time windows. In the original results, we defined the unit of observation—a market—as a therapeutic class–year. In Panel (a), the unit of observation is a therapeutic class–two year period i.e., 1996-1997 is period 1, 1998-1999 is period 2, and so on. In Panel (b), the unit of observation is a therapeutic class–five year period i.e., 1996-2000 is period 1, 2001-2005 is period 2, and so on. Robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table E.2: The market-level relationship between experimenter scale, the diversity of approaches, and success—Alternate measures of diversity and success

(a) Measuring Diversity with HHI

	HHI		Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) OLS
Average Experimenter Scale	-0.111***	-0.132***			-0.029***			0.203***
	(0.021)	(0.012)			(0.005)			(0.044)
HHI			-0.004	0.220***		-1.117***	-1.532***	
			(0.048)	(0.052)		(0.073)	(0.233)	
Market Structure Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,503	2,503	2,503	2,503	2,503	2,503	2,503	2,503
Adj R-squared	0.112	0.480	0.060	-0.026	0.063	0.198	0.163	0.133
First Stage F-statistic				124.450			124.450	

(b) Measuring Success with Drug Launch

	Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) 2SLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS
Average Experimenter Scale			-0.008*			0.069**
			(0.004)			(0.031)
Target Diversity	-0.006	-0.013**		0.098***	0.106**	
	(0.006)	(0.006)		(0.028)	(0.041)	
Market Structure Controls	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,503	2,503	2,503	2,503	2,503	2,503
Adj R-squared	0.060	0.000	0.059	0.096	0.053	0.084
First Stage F-statistic		48.557			48.557	

Notes: Panel (a) replicates the baseline results from Tables 5 and 6, where the variable Target Diversity has been replaced with HHI. A higher HHI value indicates that experiments are concentrated in a handful of approaches, and thus the market is less diverse. Panel (b) replicates the baseline results from Tables 5 and 6 where we define success as a drug development project ultimately leading to a drug product launch. Robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table E.3: Success of experimentation (1996-2018)

	Share of Pre-Clinical Success			At least 1 Pre-Clinical Success		
	(1) OLS	(2) 2SLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS
Target Diversity	-0.003	-0.046**		0.330***	0.330***	
	(0.014)	(0.016)		(0.029)	(0.035)	
Average Experimenter Scale			-0.031***			0.226***
			(0.007)			(0.039)
Market Structure Controls	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,026	2,026	2,026	2,026	2,026	2,026
Adj R-squared	0.055	-0.011	0.057	0.207	0.176	0.137
First Stage F-statistic		43.576			43.576	

Notes: This table replicates the baseline results from Table 6 with a sample including all observations up to and including 2018. Robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01

Table E.4: Controlling for target-disease discovery

	Target Diversity		Share of Pre-Clinical Success		At least 1 Pre-Clinical Success	
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS
Average Experimenter Scale	0.736***	0.724***				
	(0.094)	(0.092)				
Target Diversity			-0.017	0.040	0.309***	0.583***
			(0.014)	(0.027)	(0.020)	(0.073)
ln(Discovery)		0.032**	0.001	-0.001	0.006	-0.005
		(0.013)	(0.006)	(0.006)	(0.015)	(0.015)
Market Structure Controls	Yes	Yes	Yes	Yes	Yes	Yes
ATC-1×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	762	762	762	762	762	762
Adj R-squared	0.768	0.770	-0.006	-0.040	0.210	0.158
First Stage F-statistic				61.548		61.548

Notes: This table replicates the baseline results from columns (1)-(4) of Table 8 with the inclusion of the 2SLS estimates where we instrument for Target Diversity with Average Experimenter Scale. Robust standard errors clustered at the ATC-1 level are shown in parentheses. Significance codes: * p<.1, ** p<.05, *** p<.01