Learning About Roads Not Taken*

Daniel Schliesmann
The Wharton School

Matteo Tranchero
The Wharton School

ABSTRACT

Organizational experience is often scarce, making the ability to enrich it through generalization a central feature of learning. However, the consequences of learning in this way remain poorly understood. How does generalization shape the balance between omission and commission errors, and which organizations rely on it more? We develop a computational model that shows how generalization makes learning more efficient by spreading feedback across related alternatives, but this efficiency comes at the cost of increased omission errors. The model further predicts that the optimal degree of generalization depends on the nature of the firm's prior experience. We test these predictions using data on clinical trial failures. Consistent with the model, pharmaceutical firms reduce investment in biologically related drug targets, with the most significant declines occurring for proximate high-merit targets. The effects are larger for firms with more limited or concentrated experience, suggesting that generalization is an imperfect yet useful substitute for direct knowledge. These findings highlight both the promise and the pitfalls of generalization in settings where experience is limited.

Keywords: organizational learning, search, generalization, computational model, pharmaceutical innovation

*All authors contributed equally. We thank Justine Boudou, Dan Levinthal, Sukhun Kang, and seminar participants at the 2025 Academy of Management Annual Meeting, the Friends of Organizational Learning (FOOLs) workshop at Wharton, and the Data Innovation and AI Lab (DIAL) meeting for feedback. We are grateful to E. James Petersson for his valuable advice on our empirical setting. All errors remain our own.

1. INTRODUCTION

Learning from experience has long been recognized as foundational to organizational intelligence and strategic decision making (March and Olsen, 1975; Levitt and March, 1988; March, 1999; Argote, 2012). In classic conceptualizations of learning, an action is taken, feedback is received, and behavior is revised accordingly (Thorndike, 1913; March and Olsen, 1975; Sutton and Barto, 1998). However, an often-overlooked challenge inherent in learning from experience is that history "offers only meager samples of experience" (March et al., 1991: 1).

One setting where this scarcity of experience is evident is technological innovation. Consider, for instance, pharmaceutical drug development. In deciding which protein to pursue as a drug target, firms must navigate a vast biological search space with more than 19,000 proteins and millions of possible protein-disease combinations (Buniello et al., 2025; Tranchero, 2024). In addition to the sheer size of the search space, clinical trials are expensive — with median costs estimated at $8.6 million for Phase II and $21.4 million for Phase III trials (Martin et al., 2017) — and typically extend over several years. Strict regulatory oversight and ethical concerns about patient safety further constrain the experiments that organizations can run. As a result, pharmaceutical firms cannot rely solely on direct experience to guide their search. The scale of the opportunity space, combined with high costs, long feedback cycles, and safety concerns, severely limits the accumulation of experiential knowledge. Yet despite these challenges, firms do, in fact, learn, adapt, make informed decisions, and identify novel opportunities.

In reconciling this puzzle, an emerging body of work highlights generalization, defined as the ability to learn from "prior experience that is distinct from, but related to, one's current circumstance" (Choi and Levinthal, 2023: 1073). In this line of work, organizations take actions, receive feedback, and update their beliefs not only about the focal alternative but also about other options perceived as being related. Generalization thus allows organizations to infer the likely outcomes of untested options by transferring experience across related contexts. In pharmaceutical innovation, for example, a clinical trial on a given drug target may yield information about biologically related targets, even without direct experimentation. More broadly, this capacity to extend insights beyond narrowly accumulated experience is central to organizational learning — shaping not only drug discovery but many domains of research and development (R&D) where experi-

mentation is costly, slow, and limited.

Despite growing theoretical interest in generalization, several important questions remain. Existing treatments are largely conceptual, leaving unanswered whether and how organizations actually generalize their experience. Moreover, while generalization is often viewed as a way to compensate for limited direct experience, its consequences for search behavior are unclear. In particular, it is not yet well understood how generalization affects the likelihood of organizations committing errors of omission (overlooking promising options) and commission (pursuing poor ones). On the one hand, if related actions tend to produce similar outcomes, then generalizing feedback across alternatives may help organizations infer the merit of untested options and search more efficiently. On the other hand, learning processes often display various forms of myopia (Levitt and March, 1988; Levinthal and March, 1993). By extrapolating experience across alternatives, these biases may become especially pernicious, leading organizations to reject promising alternatives simply because they resemble past failures or to pursue poor alternatives because they resemble past successes.

These conflicting effects suggest a fundamental bias–variance tradeoff: generalization can reduce variance in the organization's beliefs by aggregating experience across related alternatives, but it may also introduce bias by propagating false positives and false negatives. This tension raises the broader question of which organizations rely more on generalization, and why. One potential explanation lies in the structure of organizations' prior experience (Eggers, 2012a). A richer and more varied experience base may allow organizations to more effectively generalize, improving their ability to extend lessons to novel alternatives (Gavetti et al., 2005; Miller and Lin, 2015). Yet abundant direct experience also reduces the need to infer from related options, making generalization less attractive if it risks introducing systematic biases. By contrast, organizations with narrower or more limited experience may depend more heavily on generalization, underscoring its role as a useful but imperfect substitute for direct knowledge.

Addressing these questions is essential for understanding how organizations navigate complex search spaces when direct experience is scarce. To explore these dynamics, we develop a computational model of organizational learning. In the model, organizations decide whether to accept or reject an alternative given limited direct experience, learn as a function of observed performance outcomes, and navigate tradeoffs between pursuing promising alternatives and avoiding costly

mistakes. We then systematically vary the nature of organizations' prior experience bases and the extent to which they generalize their experience across alternatives. This allows us to derive precise, testable predictions regarding the impact of generalizing experience on search. The model predicts that organizations that generalize will display "spatial spillovers" in learning: after a failure, they become less likely to accept not only the focal alternative but also nearby ones, with the effect weakening as the distance from the focal alternative increases. The model further predicts stronger spillovers for alternatives that are latently high performing than for those that are latently low performing, implying that generalizing negative feedback reduces commission errors while increasing omission errors. Finally, the model shows that organizations with less experience or more concentrated experience bases benefit more from generalization than those with higher levels of experience or more diffuse experience.

We then test the predictions of the model. Empirically studying how organizations learn from failure is difficult because, in many settings, it is challenging to observe credible shocks to organizational beliefs. Firms experiment selectively, outcomes are noisy, and feedback is often incomplete or ambiguous, making it hard to connect decisions to subsequent adjustments. Pharmaceutical R&D offers a rare exception (Kang, 2025). Clinical trials are large, expensive, and tightly regulated, which ensures that failures are credibly reported and widely disseminated through platforms such as the National Institutes of Health's ClinicalTrials.gov website. Because the search for effective drug targets spans millions of possible protein–disease combinations, each trial provides potential signals not only about the focal project but also about related opportunities. Failures, in particular, are salient and unanticipated events that confront firms with clear negative evidence, which is typically recognized by both the sponsoring firm and rivals (Krieger, 2021). Furthermore, we can leverage novel, granular measures of genetic distance between drug targets (Szklarczyk et al., 2025). These features make drug development an unusually clean setting for testing how organizations generalize experiential feedback.

The evidence closely mirrors the predictions of the model. Firms learn from negative feedback and sharply reduce investment in a protein–disease pair after a failed trial on that pair. These effects extend beyond the focal target: firms also scale back their innovation efforts in functionally proximate alternatives, with the magnitude of the response declining smoothly with distance in biological space. Strikingly, this spillover is strongest for neighboring high-merit tar-

gets, which firms are disproportionately likely to abandon after a related failure. This pattern reflects what Harrison and March (1984) describe as post-decision surprise: because firms typically attach higher priors to promising targets, a failure involving a related alternative generates a larger downward revision for high-merit projects than for weaker ones. The result is a systematic asymmetry in which generalization reduces commission errors by discouraging investment in low-potential drug targets, but increases omission errors by leading firms to overlook valuable opportunities. Finally, we confirm that the extent of generalization depends on firms' prior experience. Organizations with more concentrated knowledge generalize more, while those with more experience generalize less, underscoring generalization's role as a useful but imperfect substitute for direct knowledge.

Taken together, this paper makes several contributions to the literature on organizational learning and adaptation. First, it foregrounds the challenge of learning under conditions of limited experience (March et al., 1991) — a defining yet often overlooked feature of many real-world innovation environments — and shows how generalization fundamentally shapes the dynamics of organizational search. Second, it develops a computational model that formalizes generalization as a mechanism for transferring feedback across related alternatives. The model reveals a key bias–variance tradeoff. The result of this is that while generalization reduces commission errors by discouraging investment in low-quality options, it also increases omission errors by prematurely discarding high-quality ones. Importantly, this tradeoff is contingent, shifting systematically with the structure of the organization's prior experience base. These dynamics offer novel theoretical insight into when generalization is most effective and which organizations are most likely to rely on it. Third, we provide the first large-sample empirical evidence of generalization in organizational learning. Using data from pharmaceutical clinical trials coupled with novel fine-grained biological measures, we show that firms withdraw from both failed targets and related ones. Finally, by combining a computational model with this large-scale empirical analysis, we demonstrate how various conceptual and cognitive mechanisms can be measured and tested, opening the door to a more precise understanding of how organizations learn and adapt.

2. LITERATURE REVIEW

Learning can be either online, through direct experience, or offline, using existing knowledge to foresee the outcomes of untested actions (Gavetti and Levinthal, 2000). However, the literature on organizational learning is curiously asymmetric in having primarily focused on the former. For example, prior research has highlighted the phenomenon of organizational learning curves in manufacturing and other routinized tasks (Argote and Epple, 1990; Darr et al., 1995). While the contexts employed in these and other studies have varied widely, core to each of them is that learning is a function of direct experience (also known as "experiential learning" or "learning by doing"), and occurs over tasks in which the organization accumulates extensive expertise. Similarly, research grounded in the multi-armed bandit framework, a canonical representation of the exploration–exploitation tradeoff (March, 2003; Posen and Levinthal, 2012), typically assumes that learning unfolds exclusively online and over many periods (often more than 1,000) relative to a small set of alternatives (generally 2 to 10 options).

Less well understood, however, is how organizations learn when abundant direct feedback is unavailable. This kind of offline learning is particularly vital in vast and uncertain search landscapes with irreversibility in investment decisions (Adner and Levinthal, 2024). Prior work has highlighted the central role of organizations' mental models or representations (Csaszar and Levinthal, 2016; Kang, 2025) and theories (Felin and Zenger, 2009, 2017) in informing their decision-making and where they ultimately choose to direct their search and innovation efforts. However, this line of work has largely been silent on the initial development and refinement of these mental models, instead assuming them to be fixed ex ante (c.f. Gavetti and Levinthal 2000) or that organizations concurrently search through a set of multiple, pre-existing mental models to find a representation that effectively approximates reality (Csaszar and Levinthal, 2016).

Generalization from experience offers a potential explanation for how organizations develop and refine their mental models. However, this learning mechanism and its implications for organizations remain insufficiently understood. Much of the existing work is conceptual and varies widely in how generalization is operationalized, with many formulations being difficult to test empirically at scale. Prior studies have often framed generalization as a form of analogical reasoning, where learning occurs through one-to-one matches between a novel experience and some

previous domain with which the organization is familiar (Gavetti et al., 2005; Miller and Lin, 2015; Carroll and Sørensen, 2024). This process is typically one-directional: the prior experience informs the new situation, but subsequent experience with the new alternative rarely leads to a revision of the original belief. To address this limitation, later research has allowed for one-to-many, category-based learning (Martignoni et al., 2016; Choi and Levinthal, 2023), whereby alternatives are grouped into categories as a function of their perceived relatedness. In this view, organizations learn not about isolated options but about the quality of the category. Broader categories enable more extensive generalization but at the cost of reduced precision within the category (Choi and Levinthal, 2023).

Both analogical and categorical approaches, however, rely on sharp decision boundaries that treat generalization as binary. Building on more recent work, we relax this assumption and model generalization as a continuous, distance-dependent process in which its strength decays smoothly with the distance between alternatives (Schliesmann, 2025). This formulation captures a more realistic feature of organizational learning: feedback from one experience rarely transfers perfectly or not at all, but rather influences beliefs about related options in proportion to their similarity. Representing generalization in this way allows us to model how organizations generalize feedback through complex search spaces. This formulation also makes the mechanism empirically tractable, since it requires only observed performance outcomes and a measurable distance metric rather than an in-depth reconstruction of underlying matching or categorization processes. In the following section, we introduce a computational model of organizational learning to derive testable hypotheses about the nature of generalization, its implications as an adaptive mechanism, and the factors that influence how extensively organizations generalize.

3. COMPUTATIONAL MODEL

We develop a computational model to examine how organizations learn and adapt in uncertain environments. The model is designed to capture core features of adaptive decision-making, where organizations must act given limited experience, learn from performance feedback, and navigate tradeoffs between pursuing promising opportunities and avoiding costly mistakes. By varying the extent to which organizations generalize their experience across related alternatives, the model

allows us to derive precise, testable hypotheses about generalization as an adaptive mechanism in search. In particular, the model predicts its effects on organizational behavior and the conditions under which organizations are more or less likely to generalize.

3.1. MODEL STRUCTURE

We consider a model structure in which an organization responds to a stream of heterogeneous alternatives over time (Csaszar, 2013; Csaszar and Eggers, 2013; Choi and Levinthal, 2023). In each period, the organization faces a random alternative, sampled from the opportunity structure, and must decide whether to accept it and receive a stochastic payoff or reject it and receive a default payoff. Over time, the organization revises its probability of accepting different alternatives as a function of its choices and observed payoffs. To maximize performance, the organization must learn to correctly accept promising alternatives (those that generate payoffs in excess of the default reward) and reject unpromising ones (those that generate payoffs below the default reward).

3.1.1. Task Environment

The task environment consists of a set of latent alternatives, each characterized by its location in a one-dimensional trait space and by its expected performance or merit. Formally, in each period t, one alternative $i \in \{1, \ldots, N\}$ is selected at random.[1] If the organization decides to accept the alternative, the realized reward for alternative i is drawn from a Bernoulli distribution where the outcome is either a "success", generating a payoff equal to 1, or a "failure", a payoff equal to 0. Success and failure occur with respective probabilities p_i and $1 - p_i$. The underlying state of the environment is defined as a set of probabilities, $\{p_1, \ldots, p_N\}$. If the organization decides not to accept an alternative, it receives a fixed reward equal to 0.5, the mean of the underlying probability distribution.

The opportunity structure is generated by sampling values from a Gaussian Process prior

[1] In this respect, the project screening model developed here differs from an n-armed bandit model, where the organization is free to choose any alternative in each period. This choice was made to more closely map the model to the empirics, where organizations learn from both their actions and the actions of others (an exogenous choice relative to the perspective of the focal firm). While, for simplicity, we do not model multiple organizations, the results of the model can be interpreted as a representative firm in the industry responding to observed outcomes.

where each alternative occupies a discrete position in trait space $i \in \{1, \ldots, N\}$. Correlations across alternatives are governed by a radial basis function (K_{RBF}) (Wu et al., 2018; Schliesmann, 2025). Formally, the radial basis function is defined as:

$$K_{\mathrm{RBF}}(i, j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\lambda^2}\right) \tag{1}$$

where the lengthscale parameter (λ) tunes the level of autocorrelation between alternatives i and j and, by extension, the level of ruggedness in the opportunity space. For example, when $\lambda \to 0$ the opportunity space becomes increasingly uncorrelated such that each alternative is an independent and identically distributed draw from a standard normal distribution. Conversely, as $\lambda \to \infty$, the opportunity structure becomes increasingly correlated such that the autocorrelation between adjacent alternatives approaches 1, and all alternatives have the same expected performance. Once a full set of values is generated, they are then transformed into probabilities via min-max normalization over the range $[0, 1]$.[2]

3.1.2. Learning and Choice

Whether the organization accepts an alternative is informed by a process of reinforcement learning, where the probability of accepting an alternative increases following positive performance feedback and decreases following negative feedback (Thorndike, 1913; Bush and Mosteller, 1955; Lave and March, 1975; Sutton and Barto, 1998). We extend this property of learning processes to incorporate a consideration of generalization. In our formulation, organizations not only update their probability of accepting the focal alternative but also revise their likelihood of accepting neighboring ones, with the strength of this updating declining with distance in the trait space (Shepard, 1987; Wu et al., 2018, 2024; Schliesmann, 2025). The full model of probability adjustment is given by:

$$P_{j,t+1} = P_{j,t} + \phi \cdot \alpha^d \cdot (R_{i,t} - P_{j,t}) \tag{2}$$

[2]We have also assessed the robustness of the main results absent normalization. In this setting, an outcome is considered a success if its realized performance — defined as the alternative's underlying merit plus a normally distributed error term — exceeds the mean of the underlying distribution (i.e., the default payoff). Otherwise, it is considered a failure. Qualitative results are robust to this change in specification.

where $P_{j,t}$ is alternative j's acceptance probability in the current period, $R_{i,t}$ is the reward from accepting the focal alternative i in the current period, $\phi \in [0, 1]$ is the learning rate, $\alpha \in [0, 1]$ is the degree of generalization, and d is the distance between alternatives i and j. The parameter α governs how extensively the organization generalizes feedback across alternatives: when $\alpha = 0$, the organization does not engage in generalization, and the learning process collapses to the standard Bush–Mosteller fractional adjustment methodology (Bush and Mosteller, 1955; Denrell and March, 2001; Levinthal and Schliesmann, 2025); as α increases, the organization updates its acceptance probability about not only the chosen alternative but also about increasingly distant alternatives; and in the limit, when $\alpha = 1$, the organization treats all alternatives as fully interchangeable. In the main analysis, we vary the degree of generalization α to assess the implications of this mechanism on organizational adaptation. We set the baseline learning rate $\phi = 0.5$ and initialize acceptance probabilities at $P_0 = 0.5$.[3] If the organization rejects the focal alternative, no updating occurs since it receives the default payoff.

3.2. ANALYSIS

The analysis of the computational model is divided into three sections. In the first, we investigate the general behavioral patterns of organizations that generalize their experience across alternatives. Next, we investigate the implications of generalizing performance feedback across alternatives as a function of their underlying merit. In doing so, we derive several implications of generalization on the propensity of organizations to commit errors of omission and commission. Finally, we extend the model to investigate how the level and concentration of an organization's prior experience influence the extent to which it should engage in generalization.

3.2.1. Baseline Results

We first study the patterns of organizational behavior generated by the model described in Section 2.1. Thus, a set of N alternatives is sampled from a Gaussian Process with a radial basis function kernel, and a population of organizations is analyzed as responding to this environment. Follow-

[3] In addition, we have run robustness analysis across a range of ϕ values. Crucially, ϕ functions as a scaling parameter because the realized learning rate for an alternative, at a given distance, is a function of ϕ and α. The results of this analysis are reported in technical appendix A.

ing the completion of this run, a new set of N alternatives is then sampled according to the same process, and a new set of organizations is examined. This process is then repeated for 100,000 unique environments. The model is run for 100 periods and for $N = 50$ alternatives. This setting was chosen to highlight a task environment where organizations face a meaningful evaluation challenge as they are unlikely to accumulate multiple instances of direct experience with each alternative (March et al., 1991; Choi and Levinthal, 2023; Levinthal and Schliesmann, 2025). Further, the lengthscale parameter (λ) is set to an intermediate value of 1 such that the lag(1) autocorrelation across alternatives is approximately equal to 0.6.[4]

For analytical clarity, we report results following negative feedback. The mechanisms we highlight in the model operate symmetrically but not with equal magnitude: a success generates positive spillovers to related alternatives just as a failure produces negative ones; however, the effects are stronger following failure than success. The results following positive feedback are reported in Appendix A. Emphasizing failures serves to streamline exposition and highlight the learning dynamics that follow unexpected negative feedback. This choice also aligns the model with the empirical setting, where failures are typically less anticipated by firms and therefore provide cleaner identification (see discussion in subsection 5.1).

The main results are reported in Figure 1, which plots the distribution of changes in the probability of accepting an alternative in the subsequent period following a failure, as a function of its distance from the chosen alternative for a fixed level of the generalization parameter (α). In other words, the figure illustrates how the degree of generalization influences the likelihood that an organization will accept an alternative at distance x after a failure. For visual clarity, we set the α value equal to 0.8 and we present the results for various distance bins. Specifically, we plot the direct effect (distance 0), along with alternatives at distances 1-5, 6-10, 11-20, and 21-50. Turning first to the direct effect, we find, consistent with prior work on learning processes (e.g., Thorndike, 1913), that the probability of accepting an alternative declines following negative performance feedback.

[INSERT FIGURE 1 ABOUT HERE]

Turning next to the effect of negative performance feedback on unsampled alternatives (a dis-

[4]In technical appendix A, we assess the robustness of our main results to changes in the specification of each parameter.

tance greater than 0), we observe that organizations that generalize their experience display a "spillover effect" whereby the probability of subsequently accepting related alternatives declines following a failure with the sampled alternative. For example, for alternatives with absolute distances from 1 to 5, the mean reduction in the acceptance probability is approximately equal to 0.143 when α = 0.8. In contrast, the mean reduction for organizations not engaging in generalization would be equal to 0.[5] Finally, the magnitude of the spillover effect is attenuated by distance, such that the effect of failure on the change in acceptance probability is greatest for the closest alternatives. Taken together, Figure 1 suggests the following baseline hypothesis:

Hypothesis 1: *Organizations are less likely to pursue related alternatives following negative performance feedback. This generalization effect is attenuated as the distance from the focal alternative increases.*

3.2.2. Generalization and Errors of Omission and Commission

We now examine how generalization affects the likelihood of organizations' accepting alternatives that differ in their underlying merit, and in turn how it shapes the likelihood of omission and commission errors. This analysis is presented in Figure 2, which plots, as a function of the distance from the focal alternative, the change in the probability of accepting an alternative that is either latently promising or unpromising. For visual clarity, we again fix $\alpha = 0.8$ and classify alternatives as high or low merit relative to the default payoff from rejecting an alternative (0.5). Alternatives with an expected payoff greater than or equal to 0.5 are high merit and should, in principle, always be accepted, while those below 0.5 are low merit and should be rejected. Effective learning improves performance by increasing the probability of accepting high-merit options and deterring the pursuit of low-merit ones.

[INSERT FIGURE 2 ABOUT HERE]

We find that generalizing negative performance feedback has a complex effect on omission and commission errors. On the one hand, it reduces the probability that organizations accept low-

[5] Varying the degree of generalization (α) has little effect on the mean change in the likelihood of accepting the focal alternative in the subsequent period. Its influence is instead evident in the *magnitude* of spillovers: higher values of α amplify the effect of negative feedback on related alternatives. Appendix Figure A.3 reports these results in detail.

merit alternatives, with the size of this reduction declining with increased distance from the focal option. This occurs because alternatives that are close in the search space are often correlated in quality, so negative feedback on one option provides informative signals about the likely weakness of its neighbors. By spreading feedback across related choices, generalization thus helps organizations avoid repeating mistakes on similarly poor options. On the other hand, this improvement comes at a cost: the same mechanism increases the likelihood of omission errors, as promising alternatives that resemble failed ones are also discarded, leading to a systematic neglect of high-merit opportunities relative to learning without generalization.

Additionally, it is important to note the relative magnitude of the spillover effects for low- and high-merit alternatives. We find that the spillover effect is greater for high-merit alternatives than for low-merit alternatives. For example, for alternatives with a distance between 1 and 5 from the focal alternative, the reduction in acceptance probability is approximately 0.136 for low-merit alternatives and 0.15 for high-merit alternatives. This result is consistent across all distance bins and reflects the role of post-decision surprise (Harrison and March, 1984), defined as the degree to which outcomes deviate from prior expectations. As organizations gain experience, their acceptance probabilities become positively correlated with the true underlying merit, making them more inclined to pursue promising alternatives. Consequently, when negative feedback is generalized, the downward revision is larger for high-merit options, which firms had stronger priors to accept, than for low-merit ones, which were already less likely to be pursued. Taken together, this mechanism yields the following prediction:

Hypothesis 2: *Following negative performance feedback, organizations show a stronger generalization effect for high-merit alternatives than for low-merit ones, resulting in an increase in omission errors that is proportionally larger than the decrease in commission errors.*

3.2.3. Heterogeneity Analysis

We now investigate how organizations differ in their reliance on generalization. Specifically, we consider the optimal degree of generalization as a function of the level and concentration of an organization's prior experience base. As such, we extend the model structure developed in Section 3.1 to incorporate a consideration of organizations' prior experience by implementing an m

period pre-entry learning phase (Chen et al., 2018, 2024; Piezunka et al., 2022). Each organization starts the pre-entry learning phase with homogeneous initial acceptance probabilities equal to 0.5. In each pre-entry period, an alternative is selected at random and accepted by the organization with certainty. The alternative provides noisy feedback, which the organization uses to update its acceptance probability for that alternative. If the organization has a positive α value, it also generalizes this experience to neighboring alternatives. Following the completion of the m pre-entry periods, organizations will have non-homogeneous and partially informed acceptance probabilities. The simulation then commences as described in the model section.

We focus on two key parameters that characterize this pre-entry learning period, its length (m) and the concentration of experience (w). Turning first to m, if the pre-entry learning period is 0, the model collapses into the structure employed above, where organizations start with homogeneous initial acceptance probabilities. As m increases, however, the organization has a richer experience base to inform its decisions. Turning next to w, this parameter informs the probability of any given alternative being sampled in the pre-entry learning phase. Specifically, alternatives are sampled in the pre-entry period based on their distance from the previously selected alternative such that the probability of selecting alternative j given previous choice i is proportional to $(1-w)^d$ normalized by the sum of all weights:

$$P(j \mid i) = \frac{(1-w)^d}{\sum_{j=1}^{N}(1-w)^d} \tag{3}$$

where d is the distance between the previously sampled alternative i and candidate alternative j, and w is a constant bounded between [0, 1). As $w \to 1$, experience becomes increasingly concentrated such that the organization will almost exclusively sample the first alternative it selects. Conversely, as $w \to 0$, pre-entry experience becomes increasingly diffuse and, in the limit, is an unbiased random sample of alternatives, irrespective of their distance from the previously sampled alternative.

The results of this analysis on the optimal level of generalization (α*) are reported in Figure 3, with Panel A highlighting the effect of increasing the length of the pre-entry learning period (m) and Panel B highlighting the effect of increasing the level of concentration in experience (w).

We operationalize performance as the organization's final period screening accuracy (Choi and Levinthal, 2023) and consider generalization (α) values between [0,1] with step sizes of 0.05. For simplicity, we hold w constant and equal to 0 in Panel A and n constant and equal to 50 in Panel B; however, results are robust to changes in these values.[6] Turning first to the effect of pre-entry learning length (m), we find that increased experience reduces the optimal level of generalization (α^*), which declines from 0.35 when $m = 0$ to 0.05 when $m = 100$. In contrast, an increased concentration of experience (Panel B) has the opposite effect: the optimal level of generalization (α^*) increases with greater concentration, shifting from 0.2 when $w = 0$ to 0.35 when $w = 0.99$.

Taken together, generalization can be viewed as a partial substitute for both additional experience and diverse experience. Organizations with rich and varied histories can often rely directly on accumulated evidence to guide their choices, making them less dependent on inference across alternatives. By contrast, firms with narrower or more limited experience bases lack the breadth of expertise needed to evaluate new opportunities. For them, generalization becomes more central, providing a way to extrapolate from what little they know to areas they have not yet tested. In this sense, generalization enriches what would otherwise be a sparse learning environment, functioning as a compensatory mechanism that helps organizations navigate vast and uncertain search spaces when direct experience is scarce (March et al., 1991). As such, the model predicts that:

Hypothesis 3A: *Following negative performance feedback, organizations with more experience will display less pronounced generalization.*

Hypothesis 3B: *Following negative performance feedback, organizations with more concentrated experience will display more pronounced generalization.*

[INSERT FIGURE 3 ABOUT HERE]

4. EMPIRICAL SETTING

We test the model's predictions with an empirical study of how pharmaceutical firms learn from clinical trial failures. This setting allows us to observe how organizations update their R&D choices

[6] In this regard, it is worth noting that the effect of experience concentration is attenuated as the length of the pre-entry learning period (m) decreases.

in the face of negative feedback. We examine whether firms scale back their innovation efforts for drug targets that are biologically related to the failed ones, and analyze how the extent of generalization varies with the structure of their prior experience.

4.1. LEARNING FROM FAILURES IN CLINICAL DEVELOPMENT

Bringing a new drug to market is a challenging endeavor. One of the earliest and most critical decisions is selecting a drug target — usually a protein that, when modulated by a drug, can achieve a therapeutic effect (Nelson et al., 2015; Razuvayevskaya et al., 2024). With over 19,000 protein-coding genes and millions of possible protein-disease combinations, firms face a vast search space in which most candidates offer little therapeutic value (Tranchero, 2024). After a target is chosen, compounds undergo preclinical testing in animal models, followed by human clinical trials in three sequential phases of increasing cost, scale, and regulatory scrutiny. Phase I assesses safety in a small group, phase II tests efficacy in a larger sample, and phase III confirms both in a broader population over time. Despite this structured pipeline, attrition remains high: only one in ten drugs starting clinical trials ultimately receives approval from the U.S. Food and Drug Administration (Hay et al., 2014). This is because, despite extensive scientific research and testing, the actual therapeutic potential of a drug target becomes clear only when tested in large samples of human subjects.

The transparency and regulatory structure of clinical trials create a rich setting for studying organizational learning. Past empirical work on this industry shows that firms do not learn solely from their own efforts (Khanna et al., 2016; Maslach, 2016), but also observe and respond to the actions of others (Baum and Dahlin, 2007). Disclosure requirements, such as the need to register trials on the ClinicalTrials.gov platform, ensure that firms remain aware of the progress of competitors (Kao, 2025). For example, in the pursuit of treatments for Alzheimer's disease, multiple firms invested in drugs targeting the β-secretase 1 (BACE1) protein. In 2016, Eli Lilly reported the early termination of a highly anticipated phase III clinical trial targeting BACE1 due to lackluster results.[7] This failure undermined confidence in BACE1 as a drug target, prompting other firms that were pursuing it to re-evaluate their clinical investments (Krieger, 2021).

[7]The details of Eli Lilly's EXPEDITION 3 clinical trial are available online on ClinicalTrials.gov: `https://clinicaltrials.gov/study/NCT01900665`

While firms often respond to failed actions, a less-explored question is what they can learn about actions not taken. In the case of Eli Lilly's failed BACE1 trial, industry observers noted that the outcome could have implications beyond the specific compound tested (Begley, 2018). Rather than being viewed as an isolated setback, the trial was interpreted by some as a broader negative signal for functionally similar but untested targets (Garde, 2016). For example, as the name suggests, BACE2 is a close homolog of BACE1 and participates in the same biological processes (Yeap et al., 2023). From a scientific standpoint, it is reasonable to infer that the limitations evident for BACE1 might extend to BACE2 and other similarly proximate targets.

At first glance, such generalization appears sensible: failures on one alternative provide information about others that share functional similarities. Yet the organizational reality is more complicated. A long tradition of research suggests that firms often struggle to extract accurate lessons even from direct failure (Eggers, 2012a). Cognitive biases, organizational inertia, and sunk costs frequently inhibit adaptation (Ross and Staw, 1993; Tripsas and Gavetti, 2000). These challenges are possibly magnified when extrapolating to untested opportunities, where the signal is noisier and the inference less certain. Whether and how firms generalize from failure to related alternatives remains an open empirical question.

4.2. DATA AND MEASUREMENT

Human proteins constitute possible drug targets for disease treatment. We model protein–disease pairs as the set of alternatives available to firms, which together constitute the search landscape in which organizations learn from clinical trials and allocate their R&D investments. Genetic distance, measured through protein–protein interactions, captures how closely related two targets are and allows us to assess whether firms generalize feedback to nearby alternatives. To evaluate whether an organization invests in the most promising targets, we use Open Targets scores, which provide an evidence-based measure of the therapeutic potential of each protein–disease pair.

4.2.1. Clinical Trials

We use data from ClinicalTrials.gov, an online registry maintained by the National Institutes of Health (NIH) that documents ongoing and completed clinical trials. Following prior work, we fo-

cus on phase II and phase III trials, which represent significant investments in drug development (Krieger, 2021; Martin et al., 2017). In contrast, phase I trials are less consistently reported and often lack complete metadata (Kang, 2025; Kao, 2025), making them less suitable for systematic analysis. They are also less informative from a learning perspective, as early-stage failures tend to provide weaker and more ambiguous signals (Eggers, 2012b; Khanna et al., 2016). Accordingly, we construct a dataset of all phase II and III trials completed between 2001 and 2019, including information on the drug target and disease studied in each case. We stop at 2019 to avoid the confounding effects of the COVID-19 pandemic and truncation from reporting delays. Proteins are mapped to unique Gene IDs provided by the National Center for Biotechnology Information (NCBI) and diseases to Medical Subject Headings (MeSH) terms, allowing us to trace trial outcomes at the level of specific protein-disease combinations (see Appendix B for additional details on these data). To identify failures, we rely on the reported trial status and define a trial as failed if it was terminated or withdrawn prior to completion. While terminations can occur for multiple reasons, a lack of therapeutic efficacy and the emergence of side effects of the drug target are strong predictors (Razuvayevskaya et al., 2024). These early discontinuations, like the case of Eli Lilly's BACE1 trial, offer a clear setting to examine how firms respond to highly salient negative feedback.

4.2.2. Firm Patent Applications and Publications

We use data on patent applications to study how firms adapt their innovation efforts across potential targets in response to failure. Pharmaceutical firms have a well-documented tendency to patent early and frequently in the R&D process (Cohen et al., 2000), making patent applications a good real-time indicator of where they are directing their investments. Through a partnership with the European Bioinformatics Institute, we use proprietary data compiled by SciBite's TERMite software, which extracts biological entities from full USPTO patent texts between 2001 and 2019 (Tranchero, 2024). The algorithm reliably distinguishes true biological entities from casual mentions, mapping proteins and diseases to the same unique identifiers used to classify clinical trials.

To test Hypotheses 3A and 3B, we divide firms based on two dimensions of their research expertise. First, we identify higher levels of experience with a given genetic target using firms'

publication records. We draw on data from PubTator Central (Wei et al., 2024), which provides computer-annotated protein mentions for all publications indexed in PubMed. These data are matched to firms in our sample using author affiliation information, allowing us to construct a protein-level measure of expertise for each firm. A firm is classified as having higher experience if it has previously published at least once on the drug target featured in its patent application (Tranchero, 2024). Second, we measure the concentration of a firm's expertise. We compute a Herfindahl–Hirschman Index (HHI) based on the distribution of its publications across proteins, where lower values indicate a broader spread of its past research activity. A firm is classified as having a broader experience base if it exhibits a below-median HHI. This approach parallels Arts et al. (2025), who use an HHI across technology classes to assess whether firms are specialized in a narrow set of technologies. See Appendix B for more details and examples from our data.

4.2.3. Genetic Distance

To capture generalization across related genetic targets, we use a novel measure of distance based on protein-protein interactions from the STRING database (Szklarczyk et al., 2025). In this dataset, functional proximity between proteins is inferred from the frequency and strength of their interactions in human biological processes. STRING compiles and quantifies all available evidence on protein-protein associations, assigning each pair a combined confidence score that reflects the biological relevance of the interaction. This provides a biologically grounded measure of functional proximity, indicating how much feedback from one target is likely to inform others. For instance, the proteins BACE1 and BACE2 share key biological functions, which is reflected in their high interaction score (Appendix Figure B3). A key advantage of the STRING data is that it is disease-agnostic, making it particularly well-suited for studying learning in drug development, where firms often leverage the same drug target across multiple therapeutic areas. This contrasts with disease-specific approaches that rely on structural similarity, such as studying the relatedness of cancers through shared mutations (Kang, 2025).

4.2.4. Underlying Genetic Merit

To measure the underlying therapeutic potential of a drug target, we use the Open Targets score, a synthetic indicator developed by the Open Targets Platform (Buniello et al., 2025). Open Targets

is a public-private initiative that aggregates and weights all available evidence on protein-disease pairs to support clinical prioritization. It is widely regarded as the most comprehensive source of curated information on the genetic basis of human diseases. Each score reflects the strength of direct evidence linking a protein to a disease as of 2025, adjusted for the quality and reliability of the source. Recent research shows that Open Targets scores are predictive of clinical success (Razuvayevskaya et al., 2024) and are positively associated with the technological and economic value of patents targeting that protein-disease pair (Tranchero, 2024). We merge these publicly available scores with our data using Gene IDs and MeSH terms. The resulting data offer us a benchmark for evaluating the intrinsic promise of protein-disease combinations independent of firms' patenting behavior. This allows us to assess both errors of commission (when firms pursue weak alternatives) and errors of omission (when they ignore more promising ones) in the context of drug target selection.

4.3. MAPPING THE MODEL TO THE EMPIRICAL SETTING

The computational model developed in Section 3 describes organizations facing a stream of alternatives, updating their beliefs after observing each outcome, and generalizing feedback to related options in a trait space. In the empirical context of pharmaceutical R&D, these features map directly onto the search for drug targets. Protein–disease pairs represent the alternatives available to firms, while functional relatedness among proteins defines their relative position in the search space. We measure this relatedness using STRING protein–protein interaction scores, which provide an analogue to distances in the model. The merit of each protein–disease pair is captured by its Open Target score, which aggregates available scientific evidence. This corresponds to the "fitness" values in the model, with the interaction between distance and promise determining the degree of spatial autocorrelation, or landscape ruggedness. This setup allows us to test whether firms generalize feedback from failed trials to closely related proteins. Figure 4 shows the parallel between model and empirics using data on BACE1 and related drug targets for Alzheimer's.

[INSERT FIGURE 4 ABOUT HERE]

Feedback in the model arrives in the form of binary success or failure. Mimicking this feature, in our setting, pharmaceutical firms receive such feedback through the discrete outcome of clini-

cal trials. Finally, patent applications filed by firms provide a proxy for where firms are directing their investments. By relating changes in patenting activity to the timing and location of trial discontinuations, we can estimate both the direct effect of failure on the focal target and the indirect effects on genetically proximate targets.

The model further predicts that the extent of generalization is contingent on the structure of organizations' prior experience. We operationalize this prediction using two dimensions of firms' publication portfolios: their level of experience with a given protein and the concentration of their expertise across proteins. Firms without prior publications on a target represent cases of limited direct knowledge, whereas firms with highly concentrated publication activity reflect narrower domains of expertise. We conduct split-sample analyses along these dimensions to assess whether such firms exhibit stronger spillovers from clinical trial failures, as the model implies. Taken together, this empirical setting provides a unique bridge from theoretical constructs to observable organizational behavior, allowing us to test our model-derived hypotheses.

4.4. DESCRIPTIVE STATISTICS

Table 1 presents summary statistics at the level of protein-disease combinations, which constitute our primary unit of analysis. Panel A reports cross-sectional characteristics of the potential targets that pharmaceutical firms may pursue. The dataset includes 7,788,369 protein-disease pairs, constructed from 483 diseases and 16,136 human proteins featured in USPTO patents. We observe 8,683 unique phase II and III clinical trials, of which 1,578 were terminated prior to completion and are classified as failures. The clinical trials in our data provide direct information on 7,007 protein-disease pairs and generate indirect information — through varying levels of functional proximity — on an additional 2,283,284 pairs that are related to failed targets.

Panel B presents descriptive statistics for the panel dataset of protein-disease combinations observed annually from 2001 to 2019. On average, a protein-disease pair receives 0.084 patent applications per year, although the distribution is highly skewed, with some drug targets receiving more than 1,400 applications in a single year. We also split patenting activity by firms' experience bases. Roughly three-quarters of all applications originate from organizations with a higher level of experience, proxied by prior publications on the protein in question. Firms whose publication portfolios are concentrated on a narrow set of proteins account for 0.02 patents per protein-

disease pair per year, representing 20% of total patenting. Finally, Panel B compares patenting intensity by underlying scientific merit, as measured by the Open Targets score. Protein-disease pairs with a positive score receive significantly more patenting on average (0.63) than those with no score (0.052), although the latter constitute the majority of the sample.

[INSERT TABLE 1 ABOUT HERE]

5. RESULTS

We empirically estimate how firm patenting changes following a trial failure in a difference-in-difference design at the protein-disease pair level. In what follows, we discuss the research design and present the main results testing the hypotheses derived from the model.

5.1. RESEARCH DESIGN

Empirically studying how firms learn from failure presents two main challenges. The first is about measurement: knowing which actions a firm has taken and what feedback it has received. Without clear information on both, linking observed behavior to the underlying learning process is impossible. The second challenge concerns causal inference. Firms are more likely to experiment in areas where they already have expertise, such as genes they have previously (successfully) studied, which can bias estimates upward. In an ideal setting, firms would receive randomly assigned information about the therapeutic potential of specific protein-disease pairs. This would allow researchers to isolate the causal effect of such information. The impact would then be revealed by changes in patenting activity on treated protein-disease pairs relative to those left unaffected.

We approximate this ideal experiment using the staggered timing of clinical trial discontinuations. These events are publicly reported and provide shared information to all pharmaceutical firms, independent of their internal data or capabilities. Following Krieger (2021), we examine how firms respond to the failures of other companies, which helps mitigate concerns about endogeneity between firms' prior research choices and the feedback they receive. Since firms do not conduct trials with the expectation of failure, these discontinuations are largely unexpected by the

sponsoring firms and even more so by the other firms that later learn about the news. This setting offers a rare opportunity to study how organizations change their behavior in response to publicly disclosed negative outcomes.

An important strength of our approach is that it overcomes well-documented pitfalls in the empirical study of organizational learning. Traditionally, this strand of research has used the cumulative counts of past failures to predict performance changes. However, relying on cumulative variables can produce significant results even in the absence of real learning effects (Bennett and Snyder, 2017). The danger is that spurious correlations between past failures and future outcomes mechanically arise from time trends in cumulative counts. By contrast, the discontinuation of a clinical trial constitutes a discrete and externally visible shock (Krieger, 2021). This feature allows us to separate genuine behavioral adjustments from statistical artifacts. More broadly, it provides an empirical design that is well-suited to testing the model's predictions, since it yields exogenous variation in organizational feedback that can be directly traced to subsequent search behavior.

5.2. VALIDATION OF THE RESEARCH DESIGN

We first validate whether clinical trial failures act as unanticipated learning shocks by examining their impact on subsequent innovation in the same protein-disease pair. To do so, we estimate their direct learning effect with a difference-in-differences specification at the protein-disease level:

$$Y_{i,j,t} = \alpha + \beta Post_t \times ClinicalTrial_{i,j} \times Failure_{i,j} + \gamma PD_{i,j} + \delta Protein_i + \omega Disease_j + \sigma Year_t + \epsilon_{i,j,t}, \tag{4}$$

where $Y_{i,j,t}$ denotes the number of patent applications filed in year t for inventions targeting protein i and disease j. The key regressor is the interaction $Post_t \times ClinicalTrial_{i,j} \times Failure_{i,j}$, which captures whether a trial by another firm on protein-disease pair $< i, j >$ was discontinued before completion and whether the year t falls after this event. Fixed effects at the protein-disease, protein, disease, and year levels account for differences in baseline research intensity and trends across technologies and therapeutic areas. Standard errors are clustered by protein and disease. The coefficient of interest, β, isolates the effect of trial discontinuations on innovation efforts directed at the same protein-disease pair.

[INSERT TABLE 2 ABOUT HERE]

Columns 1 and 2 of Table 2 report the baseline results. While the conclusion of a clinical trial generally results in more patenting applications, there is a large and statistically significant decline following an early termination. Appendix Figure C1 further explores the stability of the difference-in-difference coefficient to alternative fixed effect structures. Interestingly, the inclusion of protein-disease pair fixed effects seems to have the largest impact on the magnitude of the coefficient. This pattern indicates that part of the raw effect reflects stable cross-sectional differences in investment intensity across protein-disease pairs, which are absorbed by the granular pair-level fixed effects, while the remaining effect captures within-pair changes in response to failure. Additional robustness checks show that results are unchanged if we focus exclusively on early terminations of Phase II clinical trials (Appendix Table C1).

The central identifying assumption of our difference-in-differences design is that, absent a trial discontinuation, patenting trends for failed pairs would have evolved in parallel to those for completed pairs. Figure 5 provides a direct test of this assumption using an event study version of Equation 4. The plot shows flat and statistically indistinguishable pre-trends and a sustained decline in patenting after the public disclosure of a failure on ClinicalTrials.gov. Post-treatment estimates stabilize at a level consistent with the average treatment effect reported in Table 2. Taken together, these results demonstrate that clinical trial failures act as well-identified negative feedback for firms, in line with prior work on organizational responses to failure (Greve, 2003; Krieger, 2021).

[INSERT FIGURE 5 ABOUT HERE]

5.3. TESTING THE THEORETICAL PREDICTIONS

We next examine whether clinical trial failures generate spillover effects on related protein–disease pairs rather than only affecting the targets directly tested. Specifically, given a clinical trial on a protein–disease pair $< i, j >$, we estimate a difference-in-differences regression for functionally related protein–disease pairs $< p, q >$:

$$Y_{p,q,t} = \alpha + \beta\ Post_t \times ClinicalTrial_{p,q}^{D(i,j)} \times Failure_{i,j} + \gamma PD_{p,q} + \delta Protein_p + \omega Disease_q + \sigma Year_t + \epsilon_{p,q,t}, \tag{5}$$

where $Y_{p,q,t}$ denotes the number of patent applications filed in year t for protein p and disease q, excluding the pair that received the clinical trial. $Clinical\ Trial_{p,q}^{D(i,j)}$ equals one if $< p, q >$ lies in distance quartile D (with $1 \leq D \leq 4$) from the failed pair $< i, j >$. As before, the indicator $Failure_{i,j}$ equals one if protein–disease pair $< i, j >$ experienced a clinical trial discontinuation. The triple interaction with $Post_t$ captures whether spillover pairs in a given quartile exhibit different post-failure dynamics relative to pairs equally distant from a successful trial. Fixed effects at the protein–disease ($\gamma PD_{p,q}$), protein ($\delta Protein_p$), disease ($\omega Disease_q$), and year ($\sigma Year_t$) levels absorb baseline heterogeneity and common trends. Standard errors are clustered by protein and by disease. The coefficient of interest, β, identifies the change in patenting for neighboring pairs within quartile D after a trial discontinuation, thereby isolating the generalization effect.

Our results show that clinical failures reverberate beyond the directly tested targets. Columns 3 and 4 of Table 2 show clear evidence of negative spillovers, with an average decline in patenting of 17.8% relative to the sample mean. Panel (a) of Figure 6 formally tests Hypothesis 1 by dividing proteins into quartiles based on their genetic distance from the nearest protein-disease pair subject to a clinical trial. The observed pattern closely mirrors the model's predictions, previously shown in Figure 1. Firms reduce innovation activity on protein-disease pairs that are functionally close to failed targets, with the effect steadily weakening across quartiles. The negative spillover becomes statistically insignificant only for the most distant group. In terms of magnitude, the generalization effect for the closest quartile is about 6% as large as the direct impact of a clinical trial failure. These results provide strong support for Hypothesis 1 and suggest that firms generalize in a distance-dependent manner from the focal alternative.[8]

[INSERT FIGURE 6 ABOUT HERE]

[8] As a falsification test, we examine whether spillovers increase when the trial signal is stronger. We use enrollment size as a proxy for signal strength, since larger trials provide more precise clinical evidence. Appendix Table C3 shows that failures of higher–enrollment trials generate larger spillovers onto related targets. This test is consistent with the intuition that firms generalize more when the feedback is clearer, and offers some additional face validity to our empirical exercise.

However, these average effects conceal important heterogeneity. Panel (b) of Figure 6 separates protein–disease pairs into high- and low-merit categories, using the Open Targets score as an indicator of scientific potential, with supporting regressions reported in Appendix Table C4. The results show that generalization reduces patenting on low-merit targets, consistent with a decline in commission errors. At the same time, it also reduces patenting on high-merit targets, implying an increase in omission errors. The effect is substantially larger for high-merit alternatives, reflecting the same asymmetric effects predicted by Hypothesis 2. Because firms tend to concentrate on a narrow set of well-known proteins (Edwards et al., 2011), a failure involving a related target generates a larger belief revision for high-merit alternatives.[9] As a result, even scientifically valuable opportunities may be abandoned, confirming a key cost of generalization driven by post-decision surprise (Harrison and March, 1984).

Finally, we examine how firms vary in their tendency to generalize from failure, depending on the level and concentration of their prior experience. Figure 7 presents these results, with supporting regressions reported in Appendix Tables C5 and C6. Panel (a) compares firms with prior scientific publications on the drug target to those without. While both groups reduce patenting as genetic distance from the failed target increases, the decline is smaller for more experienced firms, suggesting they generalize less. Panel (b) compares firms by the concentration of their prior scientific publications across proteins. Firms whose research is concentrated on a small set of proteins display stronger generalization effects, consistent with reliance on a narrow knowledge base that amplifies the need to infer from related targets. Both patterns confirm the model predictions in Figure 3 and support Hypotheses 3A and 3B. These results suggest that generalization operates as a partial substitute for experience: firms with limited or highly concentrated expertise depend on it more, but in doing so become more prone to abandoning high-potential opportunities.

[INSERT FIGURE 7 ABOUT HERE]

Taken together, the empirical evidence closely aligns with the predictions of our theoretical

[9]The difference in magnitude between the model and the empirical estimates is consistent with how belief concentration shapes generalization. In the model, the asymmetry between high- and low-merit alternatives becomes stronger when organizations hold more focused beliefs; i.e., when their priors are concentrated on a smaller set of promising opportunities. Under these conditions, failures trigger larger downward revisions for high-merit alternatives, closely matching patterns observed in the context of drug discovery.

framework. Firms learn directly from failures and generalize this feedback to biologically related opportunities in a distance-dependent way. This process reduces investment in low-quality projects, but also deters organizations from pursuing high-potential, proximate ones. This creates the asymmetry between omission and commission errors predicted by the model. The extent of generalization varies systematically with prior experience, being greater when firms' knowledge is limited or highly concentrated. Overall, the findings highlight both the value and the pitfalls of generalization, showing how it fundamentally shapes the direction of organizational search.

6. DISCUSSION AND CONCLUSION

Organizations rarely have the luxury of abundant experience. They learn, instead, in environments where histories are thin, signals are noisy, opportunities far exceed what can be directly explored due to high costs, and their choices often entail a degree of irreversibility (Adner and Levinthal, 2024). In such contexts, organizations may compensate by generalizing — extrapolating their experience from one domain to related but untested alternatives. Through generalization, organizations can augment meager experience and form beliefs about the merit of adjacent possibilities.

Yet, how organizations generalize and the consequences of this mechanism remain insufficiently understood. This paper presents theory and evidence on the conditions and organizational contingencies that make generalization more valuable to firms. Our computational model formalizes how organizations extend feedback from tested to untested alternatives, with the effect decaying smoothly with distance from the focal option. This process generates spillovers to neighboring opportunities, offering a cognitive microfoundation for the diffusion of learning observed not only across firms (Krieger, 2021), but also across activities and projects within firms (Eggers, 2012a; Zollo and Reuer, 2010). By modeling generalization as distance-dependent rather than categorical, we show how it can both accelerate adaptation under limited experience and shape systematic patterns of belief propagation within and across organizational boundaries.

We evaluate the predictions of the model in the context of pharmaceutical innovation, a domain where direct experience is highly constrained. The evidence shows how firms rely on generalization as a means to enrich sparse feedback, extrapolating lessons from known to related drug

targets. However, this mechanism is a double-edged sword. Generalization reduces variance in beliefs by aggregating across related experiences, enabling organizations to avoid overreacting to noise and to form reasonable estimates about unsampled options. At the same time, it introduces bias by increasing omission errors, as promising opportunities are abandoned when they resemble past failures. Negative spillovers are disproportionately concentrated on valuable alternatives, amplifying omission errors and reinforcing the "hot stove" effect (Denrell and March, 2001). Generalization, therefore, introduces a fundamental bias–variance tradeoff in organizational learning: it stabilizes belief updating under sparse feedback, but at the cost of systematically overlooking some valuable alternatives.

Organizations form beliefs (Posen and Levinthal, 2012), mental representations (Gavetti and Levinthal, 2000; Csaszar and Levinthal, 2016), and theories (Felin and Zenger, 2009) about the domains in which they search. These cognitive processes guide where they search, what choices they make, and the performance feedback they receive. In this regard, experience is endogenous (Denrell and March, 2001), and the challenges and myopias inherent to learning given this endogeneity have been well documented (Levinthal and March, 1993). Yet much of our theorizing assumes ample experience. In reality, as March et al. (1991) reminds us, organizations often operate with meager "samples of one or fewer". By foregrounding generalization as a partial substitute for direct experience, we reveal both its promise and its pitfalls as an adaptive mechanism, offering new insights into how organizations adapt when experience is scarce — the very conditions under which learning matters most.

REFERENCES

Adner, R. and D. A. Levinthal (2024): "Strategy Experiments in Nonexperimental Settings: Challenges of Theory, Inference, and Persuasion in Business Strategy," *Strategy Science*, 9, 311–321.

Argote, L. (2012): *Organizational learning: Creating, retaining and transferring knowledge*, Springer Science & Business Media.

Argote, L. and D. Epple (1990): "Learning curves in manufacturing," *Science*, 247, 920–924.

Arts, S., B. Cassiman, and J. Hou (2025): "Technology differentiation, product market rivalry, and M&A transactions," *Strategic Management Journal*, 46, 837–862.

Baum, J. A. and K. B. Dahlin (2007): "Aspiration performance and railroads' patterns of learning from train wrecks and crashes," *Organization Science*, 18, 368–385.

Begley, S. (2018): "What can we learn from the latest Alzheimer's drug failure?" *STAT News*.

Bennett, V. M. and J. Snyder (2017): "The empirics of learning from failure," *Strategy Science*, 2, 1–12.

Buniello, A., D. Suveges, C. Cruz-Castillo, M. B. Llinares, et al. (2025): "Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery," *Nucleic Acids Research*, 53, D1467–D1475.

Bush, R. R. and F. Mosteller (1955): *Stochastic models for learning*, John Wiley & Sons, Inc.

Carroll, G. R. and J. B. Sørensen (2024): "Strategy theory using analogy: Rationale, tools and examples," *Strategy Science*, 9, 483–498.

Chen, J. S., D. C. Croson, D. W. Elfenbein, and H. E. Posen (2018): "The impact of learning and overconfidence on entrepreneurial entry and exit," *Organization Science*, 29, 989–1009.

Chen, J. S., D. W. Elfenbein, H. E. Posen, and M. Z. Wang (2024): "Programs of experimentation and pivoting for (overconfident) entrepreneurs," *Academy of Management Review*, 49, 80–106.

Choi, J. and D. Levinthal (2023): "Wisdom in the wild: Generalization and adaptive dynamics," *Organization Science*, 34, 1073–1089.

Cohen, W. M., R. Nelson, and J. P. Walsh (2000): "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," *NBER Working Paper w7552*.

Csaszar, F. A. (2013): "An efficient frontier in organization design: Organizational structure as a determinant of exploration and exploitation," *Organization Science*, 24, 1083–1101.

Csaszar, F. A. and J. Eggers (2013): "Organizational decision making: An information aggregation view," *Management Science*, 59, 2257–2277.

Csaszar, F. A. and D. A. Levinthal (2016): "Mental representation and the discovery of new strategies," *Strategic Management Journal*, 37, 2031–2049.

Darr, E. D., L. Argote, and D. Epple (1995): "The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises," *Management Science*, 41, 1750–1762.

Denrell, J. and J. G. March (2001): "Adaptation as information restriction: The hot stove effect," *Organization Science*, 12, 523–538.

Edwards, A. M., R. Isserlin, G. D. Bader, S. V. Frye, T. M. Willson, and F. H. Yu (2011): "Too many roads not taken," *Nature*, 470, 163–165.

Eggers, J. P. (2012a): "All experience is not created equal: Learning, adapting, and focusing in product portfolio management," *Strategic Management Journal*, 33, 315–335.

——— (2012b): "Falling flat: Failed technologies and investment under uncertainty," *Administrative Science Quarterly*, 57, 47–80.

Felin, T. and T. R. Zenger (2009): "Entrepreneurs as theorists: On the origins of collective beliefs and novel strategies," *Strategic Entrepreneurship Journal*, 3, 127–146.

——— (2017): "The theory-based view: Economic actors as theorists," *Strategy Science*, 2, 258–271.

Garde, D. (2016): "A big Alzheimer's drug trial now wrapping up could offer real hope—Or crush it," *STAT News*.

Gavetti, G. and D. Levinthal (2000): "Looking forward and looking backward: Cognitive and experiential search," *Administrative Science Quarterly*, 45, 113–137.

Gavetti, G., D. A. Levinthal, and J. W. Rivkin (2005): "Strategy making in novel and complex worlds: The power of analogy," *Strategic Management Journal*, 26, 691–712.

Greve, H. R. (2003): *Organizational learning from performance feedback: A behavioral perspective on innovation and change*, Cambridge University Press.

Harrison, J. R. and J. G. March (1984): "Decision making and postdecision surprises," *Administrative Science Quarterly*, 26–42.

Hay, M., D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal (2014): "Clinical development success rates for investigational drugs," *Nature Biotechnology*, 32, 40–51.

Kang, S. (2025): "From outward to inward: Reframing search with new mapping criteria," in *Academy of Management Proceedings*, Academy of Management Valhalla, NY 10595, vol. 2025, 10129.

Kao, J. (2025): "Information disclosure and competitive dynamics: Evidence from the pharmaceutical industry," *Management Science*, 71, 5948–5970.

Khanna, R., I. Guler, and A. Nerkar (2016): "Fail often, fail big, and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry," *Academy of Management Journal*, 59, 436–459.

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

Lave, C. A. and J. G. March (1975): *An introduction to models in the social sciences*, Harper Row.

Levinthal, D. A. and J. G. March (1993): "The myopia of learning," *Strategic Management Journal*, 14, 95–112.

Levinthal, D. A. and D. Schliesmann (2025): "Cautious exploitation: Learning and search in problems of evaluation and discovery," *Organization Science*, 36, 903–917.

Levitt, B. and J. G. March (1988): "Organizational learning," *Annual Review of Sociology*, 14, 319–338.

March, J. G. (1999): *The pursuit of organizational intelligence: Decisions and learning in organizations*, Blackwell Publishers, Inc.

——— (2003): "Understanding organizational adaptation," *Society and Economy. In Central and Eastern Europe Journal of the Corvinus University of Budapest*, 25, 1–10.

March, J. G. and J. P. Olsen (1975): "The uncertainty of the past: Organizational learning under ambiguity," *European Journal of Political Research*, 3, 147–171.

March, J. G., L. S. Sproull, and M. Tamuz (1991): "Learning from samples of one or fewer," *Organization Science*, 2, 1–13.

Martignoni, D., A. Menon, and N. Siggelkow (2016): "Consequences of misspecified mental models: Contrasting effects and the role of cognitive fit," *Strategic Management Journal*, 37, 2545–2568.

Martin, L., M. Hutchens, C. Hawkins, and A. Radnov (2017): "How much do clinical trials cost?" *Nature Reviews Drug Discovery*, 16, 381–382.

Maslach, D. (2016): "Change and persistence with failed technological innovation," *Strategic Management Journal*, 37, 714–723.

Miller, K. D. and S.-J. Lin (2015): "Analogical reasoning for diagnosing strategic issues in dynamic and complex environments," *Strategic Management Journal*, 36, 2000–2020.

Nelson, M. R., H. Tipney, J. L. Painter, et al. (2015): "The support of human genetic evidence for approved drug indications," *Nature Genetics*, 47, 856–860.

Piezunka, H., V. A. Aggarwal, and H. E. Posen (2022): "The aggregation–learning trade-off," *Organization Science*, 33, 1094–1115.

Posen, H. E. and D. A. Levinthal (2012): "Chasing a moving target: Exploitation and exploration in dynamic environments," *Management Science*, 58, 587–601.

Razuvayevskaya, O., I. Lopez, I. Dunham, and D. Ochoa (2024): "Genetic factors associated with reasons for clinical trial stoppage," *Nature Genetics*, 56, 1862–1867.

Ross, J. and B. M. Staw (1993): "Organizational escalation and exit: Lessons from the Shoreham nuclear power plant," *Academy of Management Journal*, 36, 701–732.

Schliesmann, D. (2025): "The where of search," in *Academy of Management Proceedings*, Academy of Management Valhalla, NY 10595, vol. 2025, 10172.

Shepard, R. N. (1987): "Toward a universal law of generalization for psychological science," *Science*, 237, 1317–1323.

Sutton, R. S. and A. G. Barto (1998): *Reinforcement learning: An introduction*, vol. 1, MIT Press Cambridge.

Szklarczyk, D., K. Nastou, M. Koutrouli, et al. (2025): "The STRING database in 2025: Protein networks with directionality of regulation," *Nucleic Acids Research*, 53, D730–D737.

Thorndike, E. L. (1913): *The psychology of learning*, vol. 2, Teachers College, Columbia University.

Tranchero, M. (2024): "Finding diamonds in the rough: Data-driven opportunities and pharmaceutical innovation," in *Academy of Management Proceedings*, Academy of Management Valhalla, NY 10595, vol. 2024, 13751.

Tripsas, M. and G. Gavetti (2000): "Capabilities, cognition, and inertia: Evidence from digital imaging," *Strategic Management Journal*, 1147–1161.

Wei, C.-H., A. Allot, P.-T. Lai, R. Leaman, et al. (2024): "PubTator 3.0: An AI-powered literature resource for unlocking biomedical knowledge," *Nucleic Acids Research*, 52, W540–W546.

Wu, C. M., B. Meder, and E. Schulz (2024): "Unifying principles of generalization: Past, present, and future," *Annual Review of Psychology*, 76.

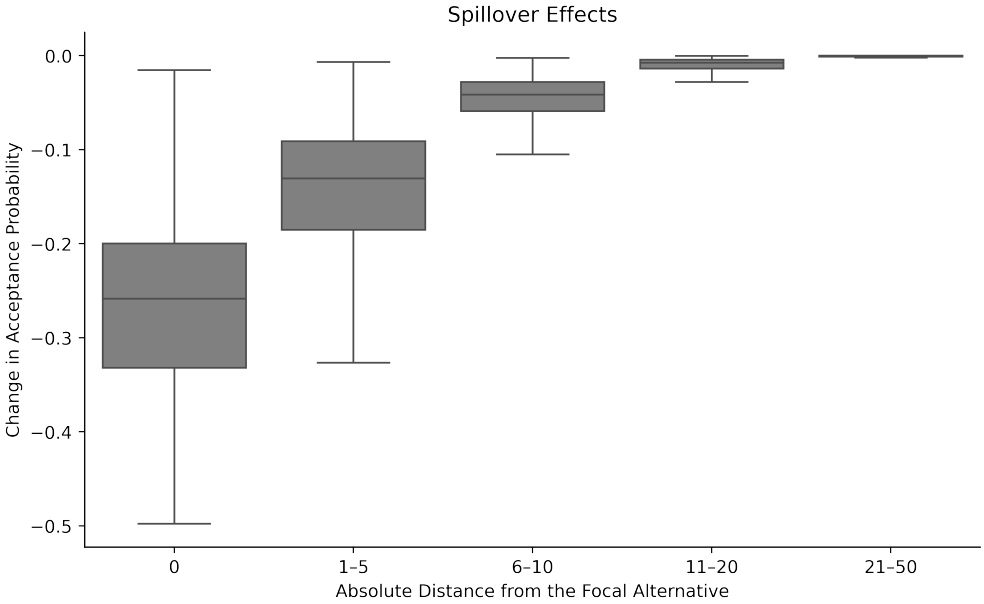
Wu, C. M., E. Schulz, M. Speekenbrink, J. D. Nelson, and B. Meder (2018): "Generalization guides human exploration in vast decision spaces," *Nature Human Behaviour*, 2, 915–924.

Yeap, Y. J., N. Kandiah, D. Nizetic, and K.-L. Lim (2023): "BACE2: A promising neuroprotective candidate for Alzheimer's disease," *Journal of Alzheimer's Disease*, 94, S159–S171.

Zollo, M. and J. J. Reuer (2010): "Experience spillovers across corporate development activities," *Organization Science*, 21, 1195–1212.

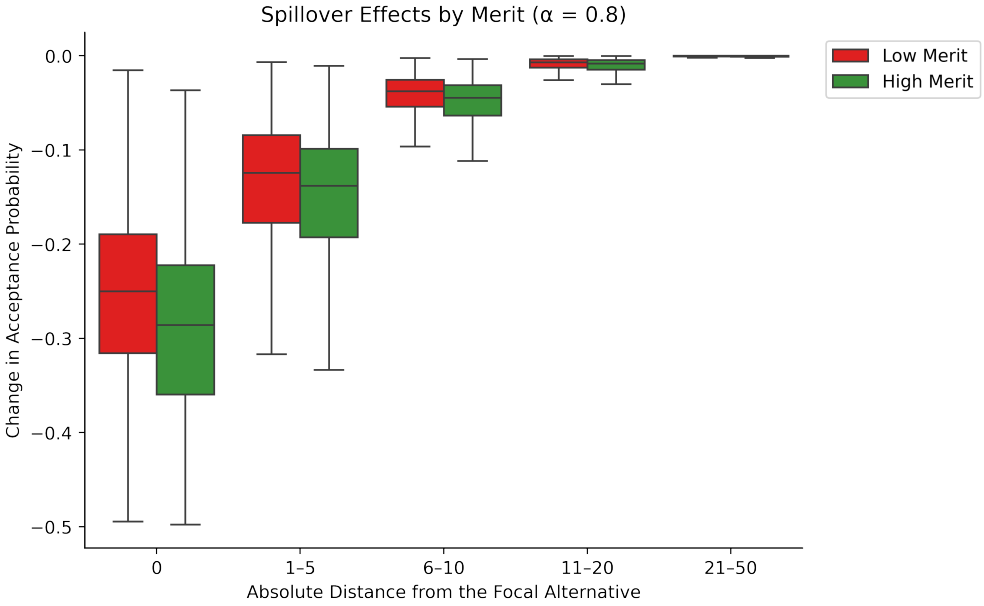
7. FIGURES AND TABLES

Figure 1: Generalization Leads to a Distance-Dependent Spillover Effect in Learning.



Note: The figure plots the effect of negative performance feedback on the probability that an organization accepts an alternative in the subsequent period, as a function of its distance from the focal alternative. For visual clarity, the degree of generalization (tuned by the parameter α) is held constant and equal to 0.8. See text for details.

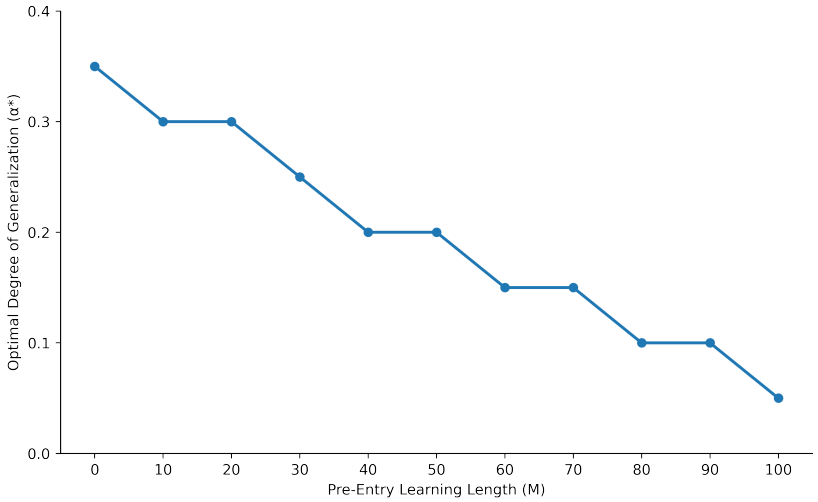
Figure 2: Generalization as a Function of the Underlying Merit of the Alternative.



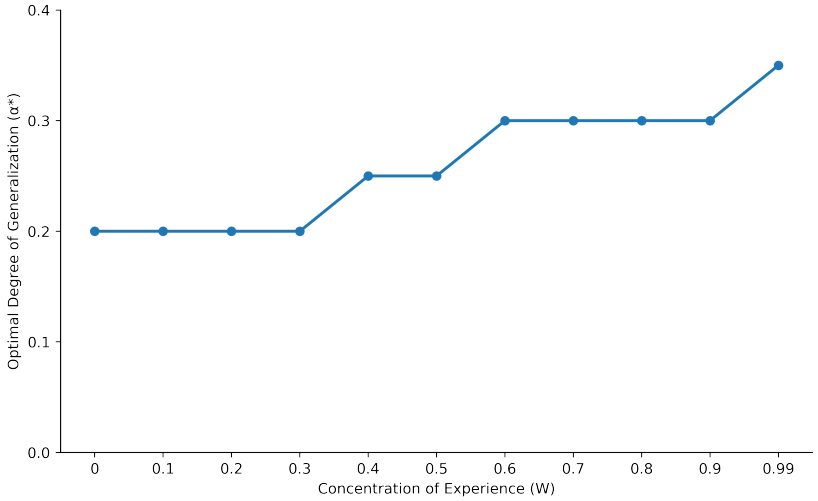
Note: The figure shows the change in the probability of accepting an alternative in the subsequent period for alternatives at varying levels of distance from the focal alternative, reported separately by different underlying merit. Low merit alternatives have an expected value less than the default payoff from rejecting an alternative (0.5). In comparison, high-merit alternatives have a value greater than or equal to the default payoff. For visual clarity, the degree of generalization (tuned by the parameter α) is held constant and equal to 0.8. See text for details.

Figure 3: Organizational Heterogeneity in the Degree of Generalization.

(a) *Optimal Degree of Generalization (α*) Varying the Level of Organizational Experience*



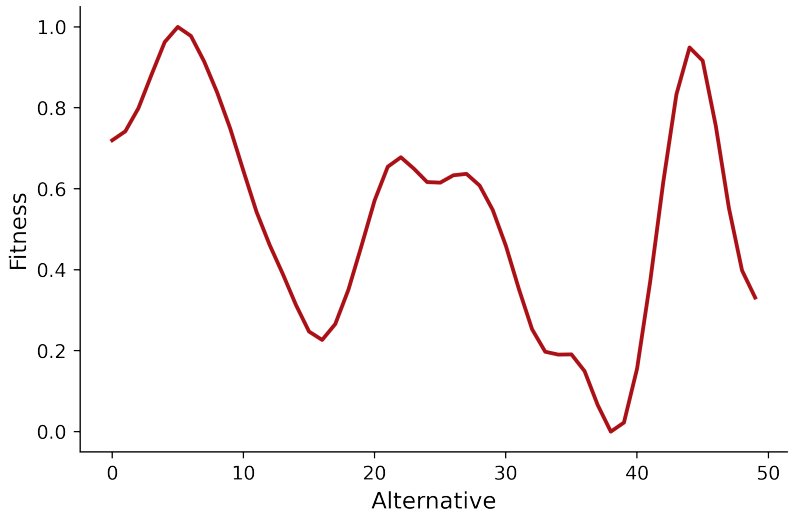
(b) *Optimal Degree of Generalization (α*) Varying the Concentration of Organizational Experience*



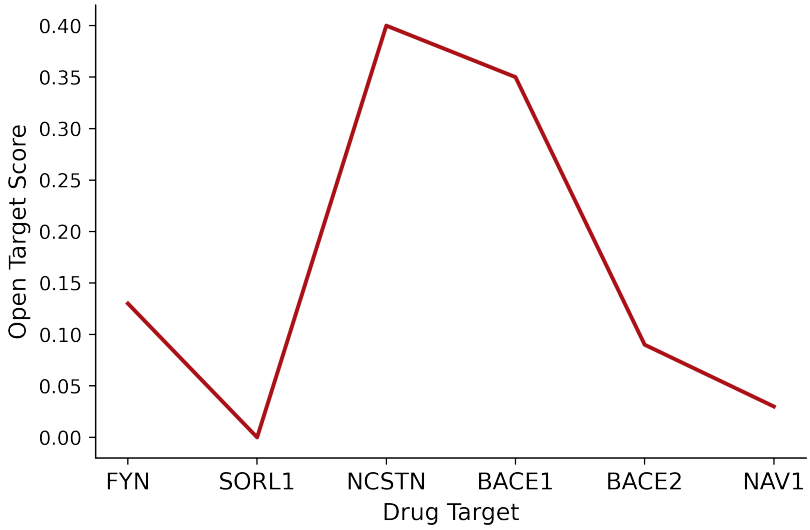
Note: The figure presents heterogeneity in the optimal degree of generalization (α*) depending on the level and concentration of organizations' pre-entry experience. Panel (a) holds the concentration of experience constant and equal to 0, such that pre-entry experience is a simple random sample of alternatives. Panel (b) holds the length of the pre-entry learning period constant and equal to 50, but changes the concentration of this experience across alternatives. See text for details.

Figure 4: Mapping the Model to the Empirical Setting.

(a) *In the Model, Organizations Search among Alternatives with Varying Fitness.*

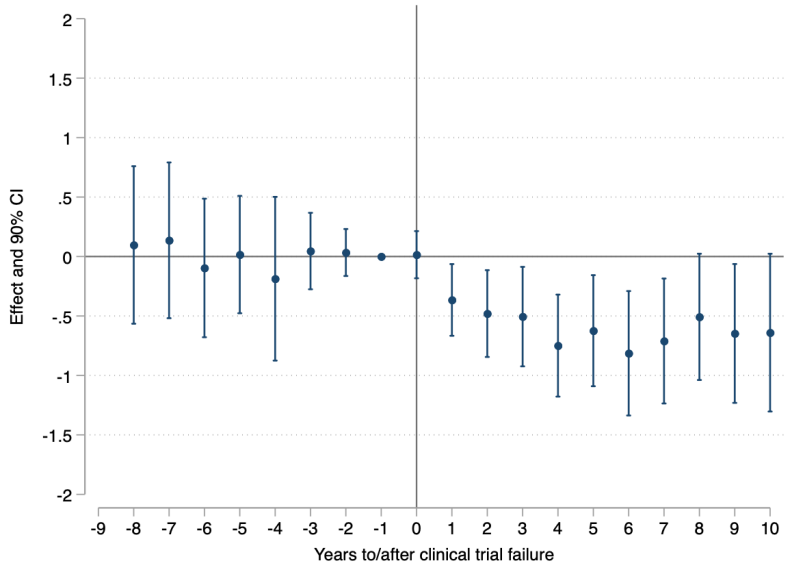


(b) *In the Data, Pharmaceutical Firms Search among Drug Targets with Varying Open Target Scores.*



Note: The figure presents an intuitive representation to map the computational model to our empirical setting. Panel (a) shows a representation of the fitness landscape where organizations search for the best alternative. Panel (b) depicts a portion of our data on drug targets for Alzheimer's disease, focusing on the BACE1 protein and drug targets closely related to it (arranged by relative distance on the X-axis). Like in the model, pharmaceutical firms search for the best drug targets among alternative proteins. See text for details.

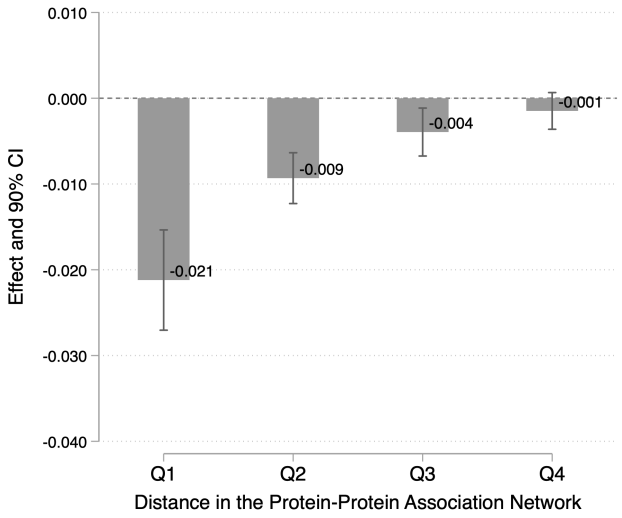
Figure 5: Patents on Focal Protein-Disease Pair Following a Clinical Failure.



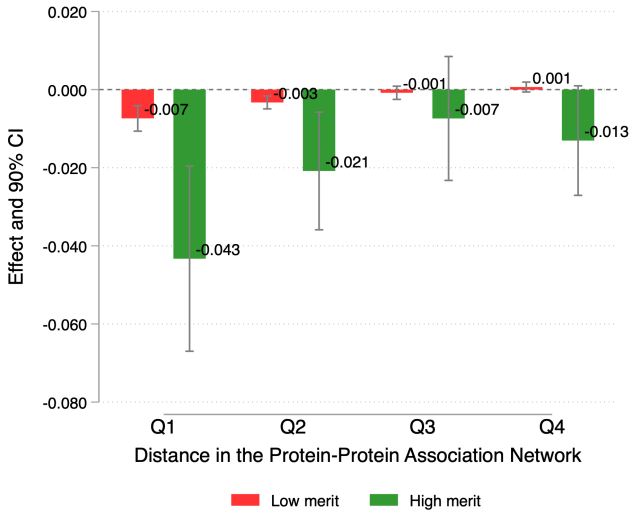
Note: The figure shows the event study coefficients estimated from the following panel OLS specification: $Y_{i,j,t} = \alpha + \sum_z \beta_z ClinicalTrial_{i,j} \times Failure_{i,j} \times 1(z) + \gamma PD_{i,j} + \delta Protein_i + \omega Disease_j + \sigma Year_t + \epsilon_{i,j,t}$. The dependent variable is the number of USPTO patent applications for innovations focusing on a specific protein-disease combination $< \ i, j \ >$ in a given year t. The chart plots values of β_z for different lags z before and after the failure of the first phase II or phase III clinical trial targeting the protein-disease pair. Regressions include protein, disease, and year fixed effects, as well as protein-disease combination fixed effects. Standard errors are clustered at the protein-disease level. See text for details.

Figure 6: Patents on Related Protein-Disease Pairs Following a Clinical Failure.

(a) *Generalization of Clinical Trial Failures.*



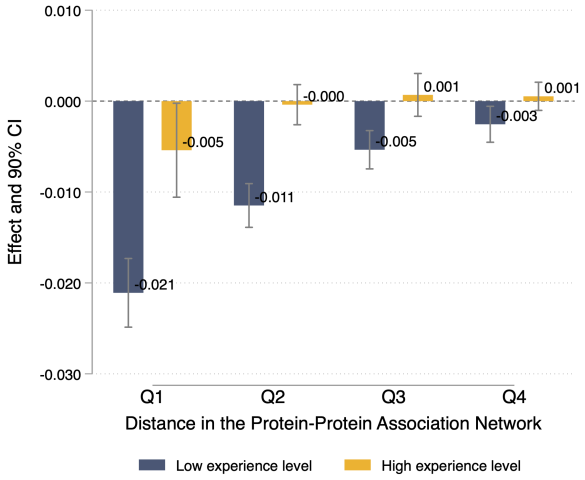
(b) *Generalization of Clinical Trial Failures by Underlying Merit.*



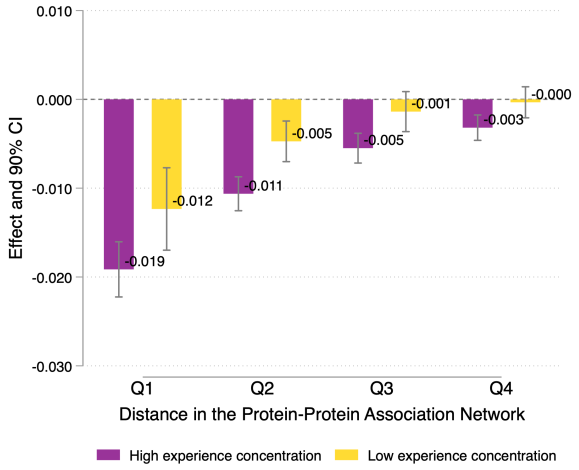
Note: The figure presents the spillover effects of clinical failures on drug targets functionally close to the failed target. Panel (a) shows the decrease in patenting for protein-disease pairs at varying levels of biological relatedness to the genetic targets of failed clinical trials. Panel (b) shows the decrease in patenting for protein-disease pairs at varying levels of biological relatedness to the genetic targets of failed clinical trials, reported separately for pairs with different underlying merit. Low-merit pairs are defined as those with an Open Targets Score of zero, while high-merit pairs have a positive score. The regressions use standardized variables to enable comparison across split-sample regressions based on Equation 5 (i.e., bars represent beta coefficients). Standard errors are clustered at the protein-disease level. See text for details.

Figure 7: Firm Heterogeneity in Patenting Behavior Following a Clinical Trial Failure.

(a) *Generalization by Level of Firm Experience.*



(b) *Generalization by Concentration of Firm Experience.*



Note: The figure presents heterogeneity in the extent to which firms generalize from clinical failures depending on their level and concentration of experience. Panel (a) shows the decrease in patenting for protein-disease pairs at varying levels of biological relatedness to the genetic targets of failed clinical trials, reported separately for firms with differing levels of genetic experience. Firms with a low level of experience are defined as those lacking prior publications on the protein. Panel (b) shows the decrease in patenting for protein-disease pairs at varying levels of biological relatedness to the genetic targets of failed clinical trials, reported separately for firms with above or below the median concentration of genetic experience. Firms with a low concentration of experience are defined as those whose publications are spread more evenly across proteins, as measured by the HHI index. The bars represent standardized beta coefficients to enable comparison across split sample regressions based on Equation 5. See text for details.

Table 1: Descriptive statistics.

	Panel A: cross sectional descriptives					
	mean	median	st d	min	max	N
Total patent applications	1.597	0	19.106	0	13900	7,788,369
Ever received clinical trial	0.000900	0	0.030	0	1	7,788,369
Ever received terminated clinical trial	0.000244	0	0.0156	0	1	7,788,369
Ever spillovers from clinical trial	0.2930	0	0.455	0	1	7,788,369
Ever spillovers from terminated clinical trial	0.0617	0	0.241	0	1	7,788,369
Average Protein-Protein Distance	85.203	0	160.909	0	999	7,788,369
Average Open Target score	0.00169	0	0.0187	0	0.908	7,788,369
	Panel B: panel descriptives					
	mean	median	st d	min	max	N
Yearly patent applications	0.0840	0	1.310	0	1465	147,979,011
. . . by firms with high level of experience	0.0233	0	0.830	0	1340	147,979,011
. . . by firms with low level of experience	0.0608	0	0.737	0	1261	147,979,011
. . . by firms with high concentration of experience	0.0172	0	0.322	0	298	147,979,011
. . . by firms with low concentration of experience	0.0669	0	1.107	0	1403	147,979,011
Yearly patent applications targeting high merit pairs	0.6310	0	4.287	0	1465	8,290,175
Yearly patent applications targeting low merit pairs	0.0520	0	0.842	0	1079	139,688,836
Post Clinical Trial (Direct)	0.0005	0	0.0216	0	1	147,979,011
Post Clinical Trial (Spillover)	0.1590	0	0.3660	0	1	147,979,011
Year	2010	2010	5.477	2001	2019	147,979,011

Note: This table lists summary statistics at the protein-disease level for 7,788,369 pairs (Panel A) and at the protein-disease-year level for a balanced panel of 147,979,011 observations (Panel B). *Total patent applications*: count of USPTO patent applications for inventions targeting a given protein-disease pair; *Ever received clinical trial*: 0/1 = 1 for protein-disease pairs that have been directly targeted by a phase II or phase III clinical trial; *Ever received terminated clinical trial*: 0/1 = 1 for protein-disease pairs that have been directly targeted by a phase II or phase III clinical trial that has been terminated; *Ever spillovers from clinical trial*: 0/1 = 1 for protein-disease pairs genetically related to targets of a phase II or phase III clinical trial; *Ever spillovers from terminated clinical trial*: 0/1 = 1 for protein-disease pairs genetically related to targets of a phase II or phase III clinical trial that has been terminated; *Average Protein-Protein Distance*: average combined score of a protein-protein interaction in the STRING database; *Average Open Target score*: average value of the Open Target score; *Yearly patent applications*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair; *Yearly patent applications by firms with high level of experience*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair filed by firms with previous publications on the protein involved; *Yearly patent applications by firms with low level of experience*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair filed by firms without previous publications on the protein involved; *Yearly patent applications by firms with high concentration of experience*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair filed by firms with previous publications concentrated on a narrow set of proteins; *Yearly patent applications by firms with low concentration of experience*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair filed by firms with previous publications spread more evenly across proteins, as measured by the HHI index; *Yearly patent applications targeting high merit pairs*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair with an Open Target score greater than zero; *Yearly patent applications targeting low merit pairs*: count of yearly USPTO patent applications for inventions targeting a given protein-disease pair with an Open Target score equal to zero; *Post Clinical Trial (Direct)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a focal protein-disease pair; *Post Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Year*= average year of observations in the panel.

Table 2: Direct and Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting.

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Direct)	0.980***	1.110***		
	(0.123)	(0.000650)		
... × Failure		-0.459**		
		(0.164)		
Post × Clinical Trial (Spillover)			0.0269***	0.0302***
			(0.000630)	(0.000713)
... × Failure				-0.0146***
				(0.00147)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	147,845,878	147,845,878

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs, while columns (3) and (4) exclude protein-disease pairs that directly received a clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Direct)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a focal protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase II or III clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

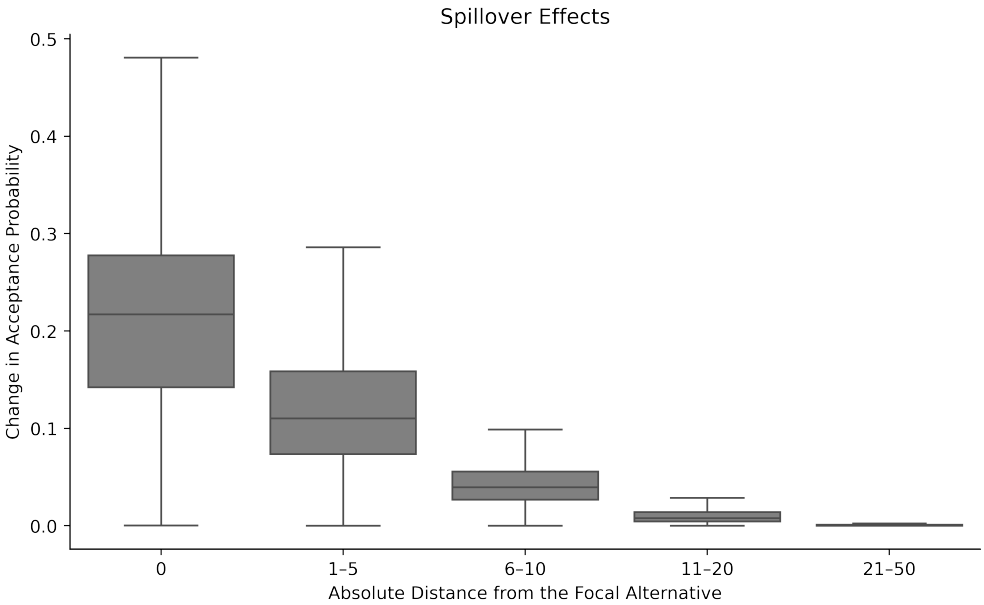
Learning About Roads Not Taken
Appendix

A. MODEL EXTENSIONS AND ROBUSTNESS CHECKS

SPILLOVER EFFECTS FOLLOWING SUCCESS:

While the main analysis highlights the effects of generalization following failure, it is also important to assess the effects of generalization following positive performance feedback. As such, in Figure A.1 we plot the change in the probability of accepting an alternative given its distance from the focal alternative. For simplicity, we hold the degree of generalization (α) constant and equal to 0.8. Ultimately, we find that generalization produces a positive spillover effect such that there is an increase in the probability of accepting similar alternatives. For example, the mean increase in the probability of accepting an alternative at a distance between 1 and 5 from the focal alternative is equal to approximately 0.12 following a success.

Figure A.1: Generalization Following Success.



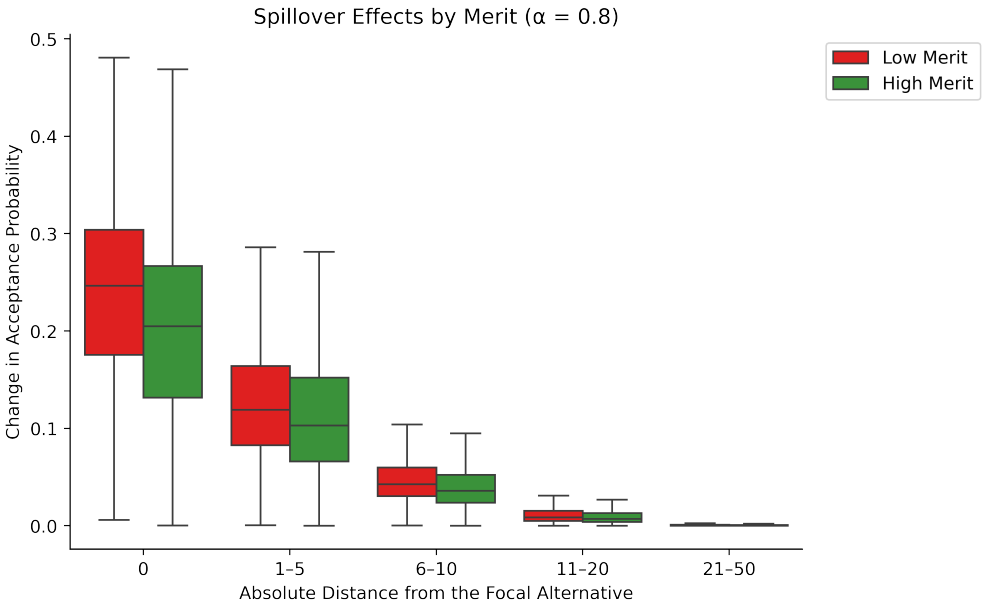
Note: The figure shows the effects of positive performance feedback on the probability of an organization accepting an alternative in the subsequent period. The effect is plotted as a function of its distance from the focal alternative. For visual clarity, the degree of generalization (tuned by the parameter α) is held constant and equal to 0.8. See text for details.

Furthermore, the main analysis highlights that the magnitude of the spillover (generalization) effect following failure is a function of the underlying merit of the alternative. We now assess the

implications of organizations' generalizing positive feedback on alternatives with varying underlying merit. This result is plotted in Figure A.2, which plots, as a function of the distance from the focal alternative, the change in the probability of accepting an alternative that is either latently promising or unpromising. For visual clarity, we fix $\alpha = 0.8$ and classify alternatives as high or low merit relative to the default payoff from rejecting an alternative (0.5). We find that generalizing positive performance feedback leads to an increase in commission errors (as the organization increases its probability of accepting a similar, low-merit alternative) and a reduction in omission errors (an increased probability of accepting similar, high-merit alternatives). Consistent with the post-decision surprise mechanism described in the main text, there is an asymmetry in the magnitude of these effects such that there is a comparatively larger increase in the probability of accepting a similar, low-merit alternative than a similar, high-merit one.

However, it is worth noting that while these results following success are symmetric to the reported results following failure, they are not of equal magnitude. Notably, the generalization effect is stronger following failure than success. For example, the average increase in the acceptance probability for the 1-5 distance bin is approximately 0.12 following success, relative to a decline of roughly 0.143 following failure. Furthermore, following success, there is an average increase in the acceptance probability of 0.113 for high-merit alternatives and an increase of 0.128 for low-merit ones. In contrast, following failure, there is a decline in the acceptance probability of 0.136 and 0.15 for low- and high-merit alternatives, respectively.

Figure A.2: Generalization Following Success by Underlying Merit of the Alternative.

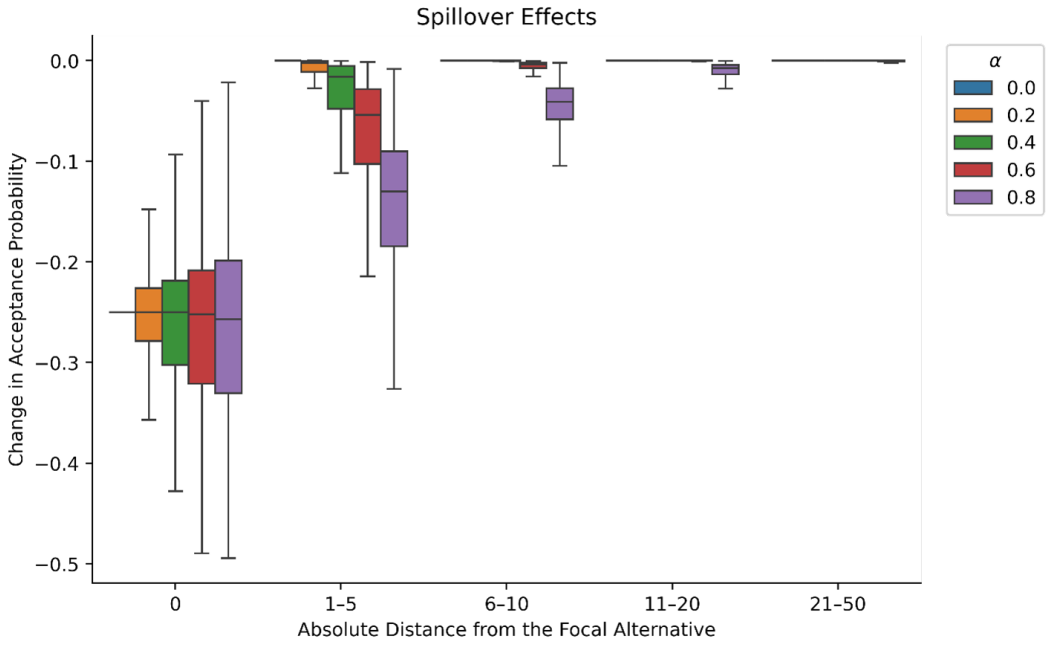


Note: The figure shows the change in the probability of accepting an alternative in the subsequent period, following positive performance feedback, for alternatives at varying levels of distance from the focal alternative, reported separately by different underlying merit. Low-merit alternatives have an expected value less than the default payoff from rejecting an alternative (0.5). In comparison, high-merit alternatives have a value greater than or equal to the default payoff. For visual clarity, the degree of generalization (tuned by the parameter α) is held constant and equal to 0.8. See text for details.

SPILLOVER EFFECTS VARYING GENERALIZATION:

The main analysis highlights that generalization reduces the probability of accepting both the focal alternative and related alternatives following failure. Further, while Figure 1 highlights that the magnitude of this reduction is a function of the distance between the accepted alternative and the related alternative, it is also important to note how the extent of generalization (α) shapes the magnitude of this effect. In Figure A.3 we plot the change in acceptance probability of alternatives varying in their distance from the focal alternative for alternative levels of generalization (α). Ultimately, while the extent of generalization does not dramatically affect the magnitude of the direct effect (the change in acceptance probability for the focal alternative), it does affect the magnitude of the spillover effect. For example, the mean reduction in the probability of accepting an alternative at a distance between 1 and 5 is equal to approximately 0.143 when α = 0.8, but is reduced to approximately 0.07 when α = 0.6.

Figure A.3: Spillover Effects Varying the Degree of Generalization (α).



Note: The figure shows the effects of negative performance feedback on the probability of an organization accepting an alternative in the subsequent period. The effect is plotted as a function of its distance from the focal alternative and the degree to which the organization generalizes its experience (tuned by the parameter α). See text for details.

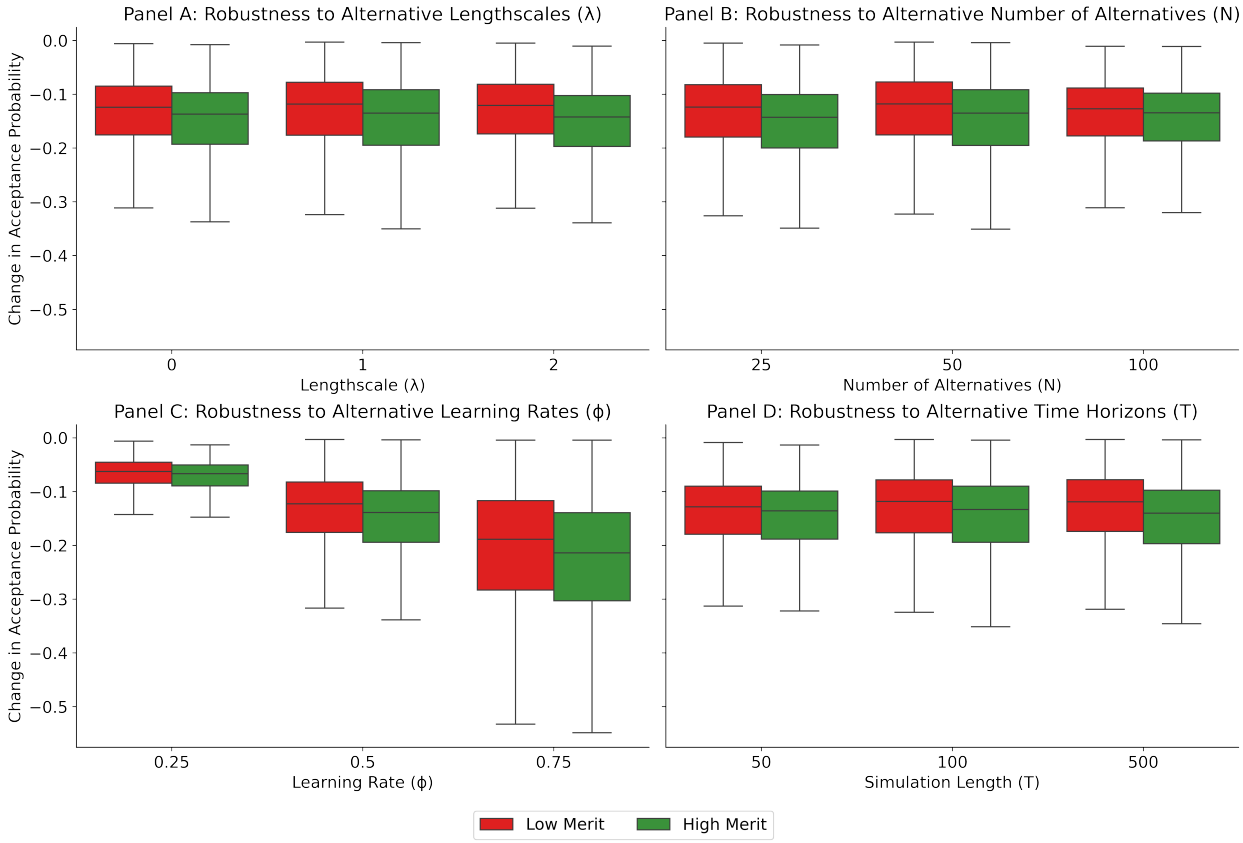
SPILLOVER EFFECTS BY MERIT:

The main analysis highlights that generalization produces a spillover effect and that the magnitude of this spillover effect is a function of the underlying merit of the alternative. We now assess the robustness of these findings across a wide range of alternative parameter settings. As such, Figure A.4 plots the distribution of changes in the probability of accepting either a high- or low-merit alternative in the subsequent period following failure. As with the main analysis, we classify alternatives as high or low merit relative to the default payoff the organization would receive from rejecting an alternative (0.5). More specifically, high-merit alternatives have an expected value greater than or equal to 0.5, while low-merit alternatives have an expected value below 0.5. For visual clarity, we plot the spillover effect only for those alternatives with an absolute distance of 1-5 from the focal alternative and hold the generalization parameter α constant and equal to 0.8.

Panel A plots these results, varying the lengthscale parameter (λ), thereby allowing us to assess the sensitivity of our findings to the underlying level of spatial correlation in the task environment. Specifically, we highlight λ values approaching 0 (a setting where each alternative is an iid draw from the generating distribution), the baseline setting of λ = 1 (which results in the spatial correlation across adjacent alternatives being approximately equal to 0.6), and λ = 2, a setting where alternatives are very highly correlated (with the spatial correlation across adjacent alternatives being approximately 0.9). Ultimately, while the lengthscale parameter affects the optimal degree of generalization (see Figures A.5 and A.6), it has only a minimal effect on the presence of this spillover effect, the magnitude of the effect, and that the spillover effect is greater for high-merit than low-merit alternatives.

Panel B plots these results, varying the size of the search space between a setting with fewer alternatives (25 alternatives instead of the 50 alternatives in the baseline setting) and a setting where the number of alternatives is of the same scale as the number of periods (100). Ultimately, over the range of values considered, the number of alternatives has only a minimal effect on the presence of this spillover effect, the magnitude of the effect, and the relative magnitude of this spillover effect for high- and low-merit alternatives. Turning to Panel C, we assess the effect of tuning the magnitude of the learning rate (ϕ). Specifically, when ϕ is decreased from 0.5 to 0.25 (a slower learning rate), the magnitude of the spillover effect is attenuated, whereas increasing ϕ to 0.75 has the opposite effect. The dependence of the spillover magnitude on the learning rate is expected, given that the effective updating rate at any distance is jointly governed by ϕ and α. Notably, while ϕ shifts the magnitude of this spillover effect, we continue to observe that the median change in the probability of accepting an alternative following negative performance feedback is greater for high-merit than low-merit alternatives. Finally, Panel D assesses the sensitivity of our results to changes in the number of periods (T). The results are highly consistent for both shorter (50 periods) and longer (500 periods) time horizons.

Figure A.4: Spillover Effects at Distance 1-5 by Underlying Merit.



Note: The figure shows the change in the probability of accepting an alternative in the subsequent period for alternatives with an absolute distance from the focal alternative between 1 and 5, reported separately for low- and high-merit alternatives. Low-merit alternatives are defined as those with an expected value less than the default payoff from rejecting an alternative (0.5), while high-merit alternatives have a value greater than or equal to the default payoff. For visual clarity, the degree of generalization (tuned by the parameter α) is held constant and equal to 0.8. Each panel varies one parameter of the model while holding all other parameters constant and equal to the values employed in the main analysis. See text for details.

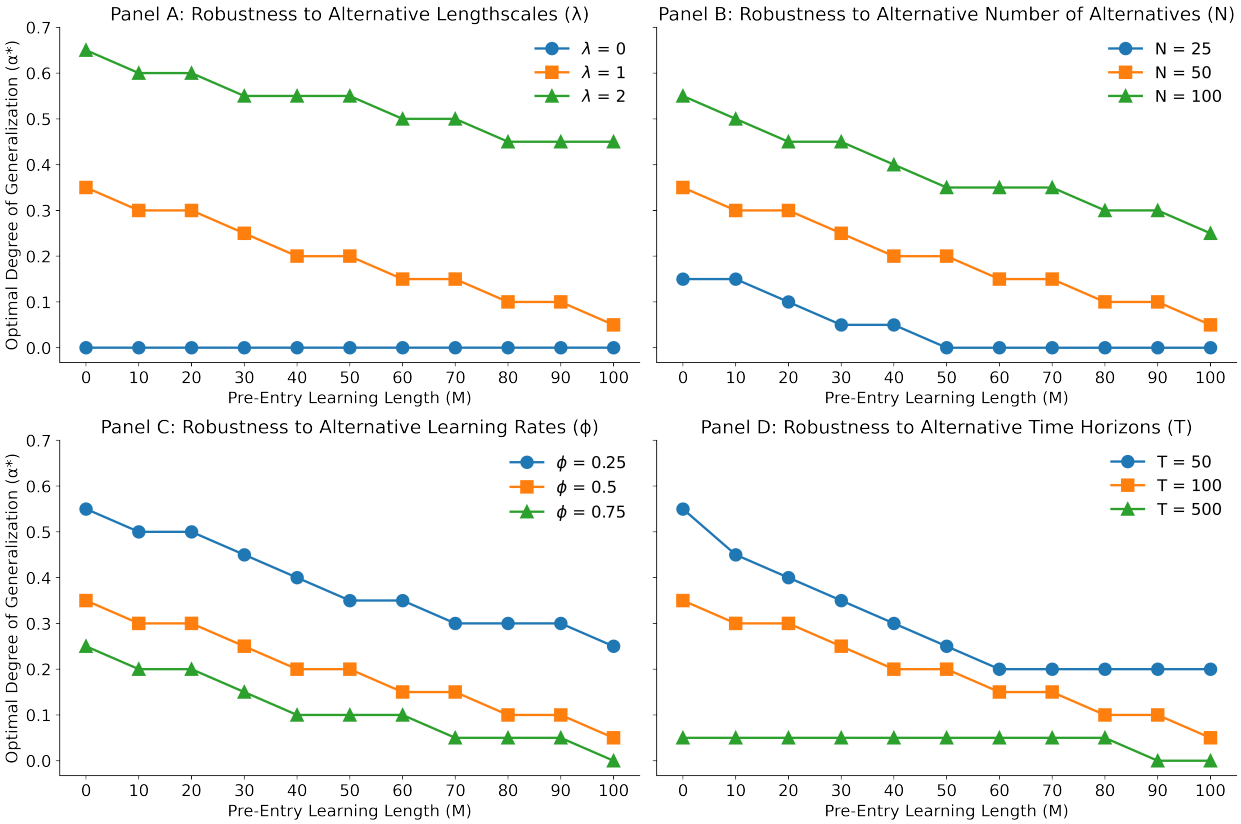
LEVEL OF INITIAL EXPERIENCE:

While the main analysis highlights that the optimal degree of generalization declines as a function of the level of the organization's prior experience base, it is important to assess the robustness of these findings across a wide range of parameter settings. This analysis is presented in Figure A.5 which plots how the optimal level of generalization shifts as a function of the length of the pre-entry learning period m. Panel A highlights this result while varying the lengthscale parameter (λ). In doing so, we assess the sensitivity of our findings to the underlying level of spatial correlation in the task environment. We find that in settings absent spatial correlation across alternatives (λ = 0), the optimal level of generalization (α*) is consistent and equal to 0, such that organizational adaptation is best served by not engaging in generalization. Conversely, increasing λ to 2 has the effect of shifting the experience-α* curve upwards such that, over the range of m values considered, adaptation is best served by a higher level of generalization than the baseline setting of λ = 1. Further, in this setting, the finding that the optimal level of generalization declines as a function of experience is maintained.

Panel B plots these results while varying the size of the search space from 25 alternatives (a smaller number of alternatives relative to the main analysis) to a setting where the number of alternatives is of the same scale as the number of periods (100). Ultimately, we find, over the range of the number of alternatives considered, that increased experience continues to have the effect of reducing the optimal level of generalization. Furthermore, changing the number of alternatives shifts the experience-α* curve up and down such that in settings with fewer alternatives, organizational adaptation is best served by, on average, a lower level of generalization, whereas in settings with a greater number of alternatives, optimal performance is associated with, on average, higher levels of generalization. Turning to Panel C, we assess the effect of tuning the magnitude of the learning rate (ϕ). Specifically, when ϕ is decreased from 0.5 to 0.25, on average, the optimal level of generalization increases (the experience-α* curve is shifted upwards), whereas increasing ϕ has the opposite effect. Crucially, across the range of ϕ values considered, we continue to observe that increasing experience has the effect of reducing the level of generalization associated with the greatest level of performance.

Finally, Panel D assesses the sensitivity of our results to changes in the length of the simulation (T). We find that when T is decreased from 100 to 50, on average, the optimal level of generalization increases (the experience-α* curve is shifted upwards), whereas increasing T has the opposite effect. That shifting the time horizon under consideration has this effect on the level of the experience-α* curve should not be surprising, as increasing the length of the simulation effectively shifts the level of experience from which the organization has to draw on in deciding whether to accept or reject an alternative. Ultimately, while the length of the simulation shifts the level of the experience-α* curve, the finding that the optimal degree of generalization declines as a function of experience is maintained across the range of time periods considered.

Figure A.5: Optimal Degree of Generalization (α*) by Experience Level



Note: The figure presents heterogeneity in the optimal degree of generalization (α*) depending on the level of organizations' pre-entry experience. The concentration of experience (w) is held constant and equal to 0, such that pre-entry experience is a simple random sample of alternatives. Each Panel varies one parameter of the model while holding all other parameters constant and equal to the values employed in the main analysis. See text for details.

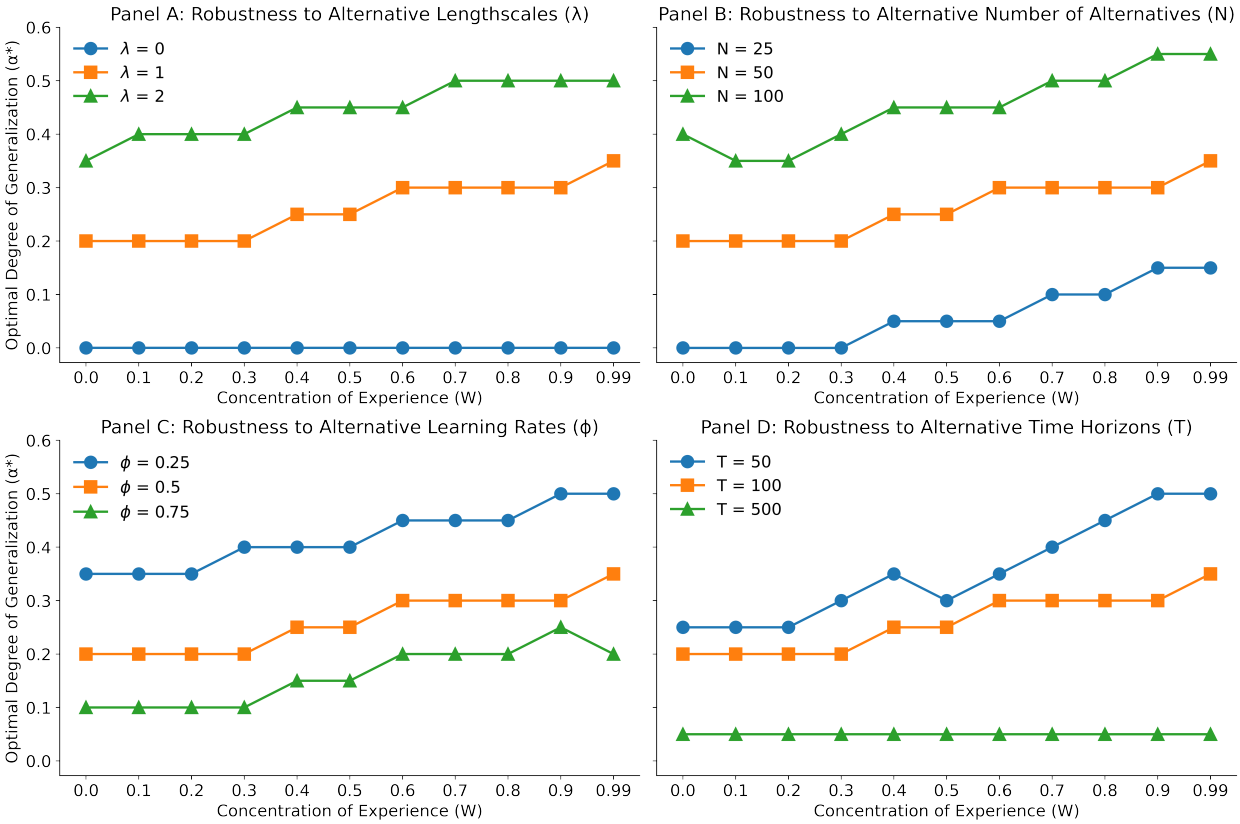
CONCENTRATION OF INITIAL EXPERIENCE:

While the main analysis highlights that the optimal degree of generalization increases as a function of the level of concentration in the organization's prior experience base, it is important to assess the robustness of these findings across a wide range of parameter settings. This analysis is presented in Figure A.6 which plots how the optimal level of generalization shifts as a function of the concentration of the organization's pre-entry experience w. Panel A highlights this result while varying the lengthscale parameter (λ). In doing so, we assess the sensitivity of our findings to the underlying level of spatial correlation in the task environment and find that in settings absent spatial correlation across alternatives (λ = 0), the optimal level of generalization (α*) is consistent and equal to 0, such that organizational adaptation is best served by not engaging in generalization. Conversely, increasing λ to 2 has the effect of shifting the concentration-α* curve upwards such that adaptation is, on average, best served by a higher level of generalization than in the baseline setting of λ = 1. Further, the finding that the optimal level of generalization increases as a function of the level of concentration in experience is maintained.

Panel B plots these results while varying the size of the search space from a setting with 25 alternatives (a smaller number of alternatives than the 50 alternatives employed in the main analysis) to a setting where the number of alternatives is of the same scale as the number of periods (100). We find, over the range of the number of alternatives considered, that increasingly concentrated experience continues to have the effect of increasing the optimal level of generalization. Additionally, changing the number of alternatives shifts the concentration-α* curve up (down) such that in settings with fewer (more) alternatives, organizational adaptation is best served by, on average, a lower (higher) level of generalization. Turning to Panel C, we assess the effect of tuning the magnitude of the learning rate (ϕ). Specifically, when ϕ is decreased from 0.5 to 0.25, on average, the optimal level of generalization increases (the concentration-α* curve is shifted upwards), whereas increasing ϕ to 0.75 has the opposite effect. Crucially, across the range of ϕ values considered, we continue to observe that increasing the concentration of experience reduces the level of generalization associated with the greatest level of performance.

Finally, Panel D assesses the sensitivity of these results to changes in the length of the simulation (T). We find that when T is decreased from 100 to 50, on average, the optimal level of generalization increases (the concentration-α* curve is shifted upwards), whereas increasing T has the opposite effect. Further, while the length of the simulation shifts the level of the concentration-α* curve, the finding that the optimal degree of generalization increases as a function of the concentration of experience is consistent across the range of time periods considered.

Figure A.6: Optimal Degree of Generalization (α*) by Experience Concentration



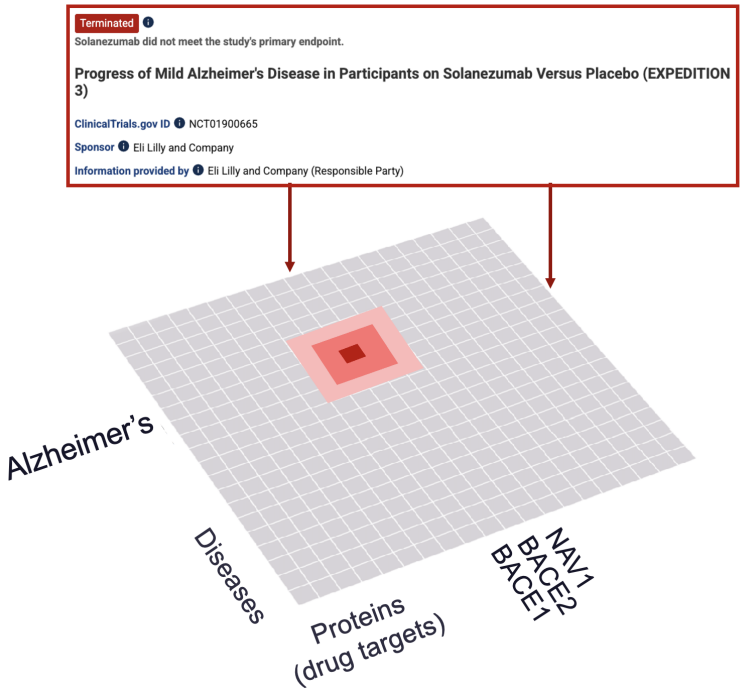
Note: The figure presents heterogeneity in the optimal degree of generalization (α*) depending on the concentration of organizations' pre-entry experience. The level of experience (m) is held constant and equal to 50. Each Panel varies one parameter of the model while holding all other parameters constant and equal to the values employed in the main analysis. See text for details.

B. MEASUREMENT DETAILS

PHARMACEUTICAL FIRMS SEARCH IN A PROTEIN-DISEASE LANDSCAPE:

Our measurement strategy builds on the idea that firms search for valuable drug targets within a vast landscape of protein–disease combinations. In our data, this landscape spans 16,136 human proteins and 483 diseases, generating more than 7.7 million possible pairs. The number of proteins is slightly below the approximately 19,000 found in the human body because we restrict attention to those mentioned at least once in a patent; however, results are robust to including proteins with no recorded R&D activity. Each pair represents a potential direction for firms' R&D, and the distances among them define the topology of the search landscape commonly studied in the organizational literature.

Figure B1: Clinical Trial Outcomes Provide Information about the Tested Protein-Disease Pair and Neighboring Ones.



Note: The figure provides a stylized illustration of how a clinical trial, such as Eli Lilly's EXPEDITION 3 trial on BACE1 inhibitors for Alzheimer's disease, conveys information about a broader region of the firm's search space. The intensity of the red shading indicates the strength of the information transmitted at varying distances from BACE1, with darker shades representing stronger signals. See text for details.

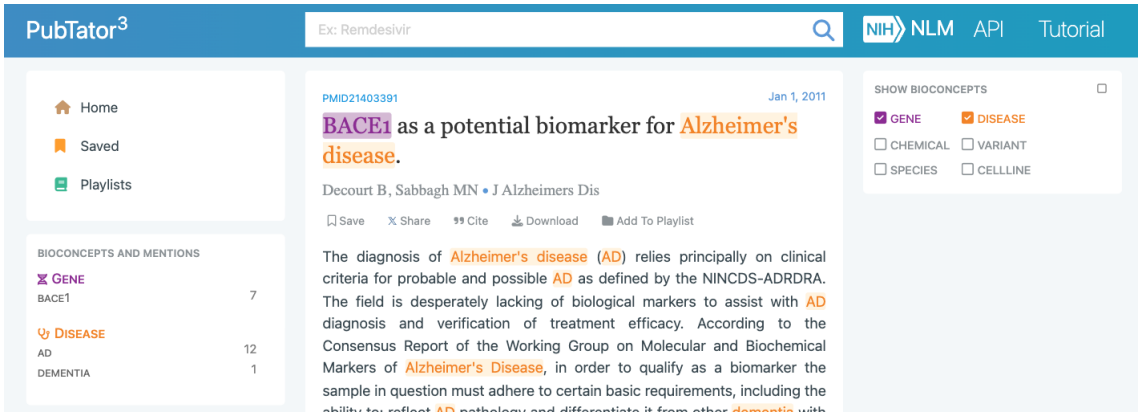
Mapping clinical trial outcomes onto this landscape is the first step in our analysis. A trial that tests a specific protein–disease pair corresponds to an experiment at a single cell in the landscape. The outcome of that experiment provides feedback not only about the tested pair but also about nearby regions. This structure allows us to examine both direct learning on the focal protein–disease pair and spillovers to related pairs at varying biological distances. Figure B1 provides a visual illustration. Empirically, we link trials to the landscape using information on the

disease conditions studied in each trial, uniquely identified by MeSH IDs, and the proteins targeted by the intervention, identified by NCBI's Gene IDs.[10] Previous studies have used this approach to study Phase I trials (Kang, 2025); here, we focus on Phases II and III, where data coverage and reliability are substantially higher (Kao, 2025).

We then trace firm investments using USPTO patent applications. In collaboration with the European Bioinformatics Institute, we use data compiled with SciBite's TERMite software, which extracts biological entities from full patent texts and links them to standardized identifiers (Gene IDs for proteins and MeSH terms for diseases). TERMite is a proprietary tool specifically designed for disambiguating biomedical text. These data, previously used and validated by Tranchero (2024), have been shown to be highly accurate. This approach allows us to position each patent application within the same landscape as the clinical trials. In turn, we can observe whether a firm patents the exact protein–disease pair tested in a trial or a pair nearby in the landscape. Because project-level R&D spending data are rarely available, patent applications provide a real-time proxy for where firms allocate resources early in the innovation process.

Finally, we measure firms' prior knowledge using scientific publications from PubMed, processed through open data released by PubTator3 (Wei et al., 2024). Each publication is tagged with the same disease and protein identifiers mentioned above, as illustrated in Figure B2. We link these publications to firms using affiliation data from Dimensions, allowing us to capture each firm's accumulated expertise with specific drug targets prior to patenting. Combining this information with our clinical and patent data enables us to examine how the composition of a firm's knowledge base shapes its response to trial outcomes. In particular, we test whether firms with deeper and broader genetic experience generalize less from failures, while those with narrower or more limited experience generalize more.

Figure B2: Pubtator3 Data Extract the Proteins and Diseases Studied in Each Published Paper.



Note: The figure shows how PubTator3 annotated scientific articles by extracting the proteins and diseases mentioned in their title and abstract. See text for details.

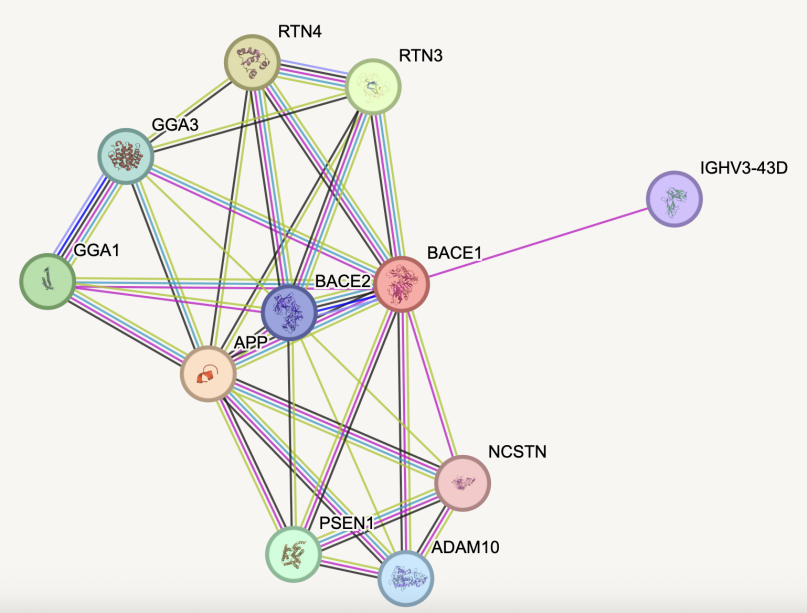
[10]We leverage the fact that proteins are coded by the homonym gene and thus use unique Gene IDs for our data matching.

DISTANCE AND FITNESS OF IN THE PROTEIN-DISEASE LANDSCAPE:

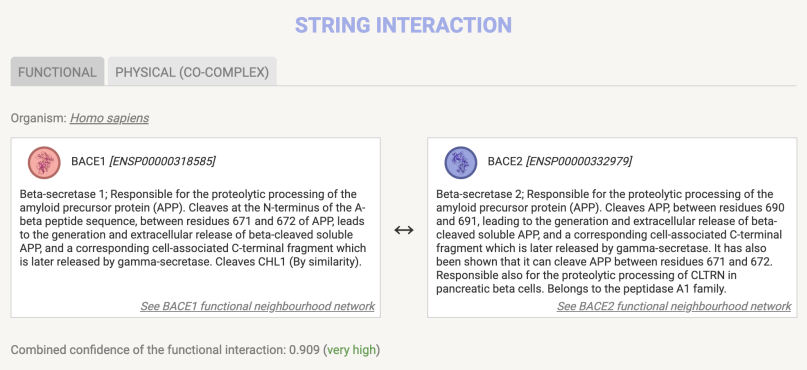
To study generalization empirically, we need a measure of distance between genetic targets. The model predicts that firms generalize feedback from one alternative to others in proportion to their similarity, with the effect decaying as distance increases. Capturing this mechanism in the data requires a biologically meaningful measure of proximity between proteins that varies continuously rather than relying on discrete categories or binary overlaps.

Figure B3: Examples of Functional Distance between Human Proteins in the STRING Database.

(a) *Network of protein-protein interactions for the BACE1 protein*



(b) *Functional distance between the BACE1 and BACE2 proteins*



Note: Panel (a) plots the protein-protein associations network centered on the BACE1 protein according to the data in STRING. Panel (b) reports the details of the interaction between BACE1 and BACE2. See text for details.

We measure this proximity using STRING (`https://string-db.org/`), a database that integrates known and predicted protein–protein associations from five evidence channels: genomic context predictions, high-throughput experiments, conserved co-expression, automated text mining, and curated pathway databases (Szklarczyk et al., 2025). STRING combines these sources

into a single confidence score for each protein-protein link, which we use as a continuous measure of functional relatedness. Higher scores denote closer proteins in the biological landscape and thus greater potential for knowledge spillovers. Our conversations with chemistry researchers confirmed that STRING is widely used in both medicinal chemistry and molecular biology. In drug discovery, researchers use it to interpret large genomic data after perturbing a target and to identify affected pathways. In experimental biology, it helps verify whether observed protein interactions align with established functional links. These applications confirm that STRING captures the kind of biological proximity along which firms are likely to generalize feedback.

Figure B3 shows an example of how STRING captures functional proximity among proteins. Panel (a) shows the network of associations centered on the BACE1 protein, while Panel (b) highlights its strong functional link with BACE2. As their names suggest, BACE1 and BACE2 are homologous proteins belonging to the same gene family, formed by duplication of a common ancestral gene and sharing similar biochemical functions. BACE2 shares approximately 64% amino acid sequence similarity with BACE1 and participates in the same biological processes (Yeap et al., 2023). To validate our measure, we verified that proteins from the same family exhibit higher STRING proximity scores, on average 14% greater than for unrelated proteins. Importantly, STRING captures functional rather than purely structural similarity, which is appropriate for our context since proteins can be structurally similar yet play distinct biological roles, just as structurally dissimilar proteins can serve as functional substitutes.

As noted previously, we measure the "fitness" of each protein–disease pair using the Open Targets score, which aggregates genetic and biomedical evidence linking a protein to a disease. The Open Targets Platform is a public–private partnership that compiles all available evidence on gene–disease associations and summarizes it in a synthetic score (Buniello et al., 2025). Each source of evidence is weighted according to a scoring framework, and the resulting values are harmonized to standardized identifiers for proteins (Gene IDs) and diseases (MeSH terms). These data are openly available online through the Open Targets Platform (`https://platform.opentargets.org/`) and were merged with our dataset. The scores provide the most comprehensive and systematic measure of the strength of genetic evidence currently available, and they are highly predictive of future clinical trial success (Razuvayevskaya et al., 2024). We therefore use the Open Targets score as an evidence-based proxy for the underlying therapeutic potential of each protein–disease pair.

Figure 4 in the main text illustrates how our setup gives rise to a fitness landscape, shown there for the case of BACE1 and its neighboring proteins in relation to Alzheimer's disease. Relative to BACE1, the STRING interaction scores (our measure of distance) for selected neighboring proteins are as follows: BACE2 = 0.909, NCSTN = 0.878, SORL1 = 0.737, NAV1 = 0.666, and FYN = 0.438. Each of these proteins also has an Open Targets score that captures its biological and therapeutic potential for Alzheimer's disease: BACE1 = 0.35, BACE2 = 0.09, FYN = 0.13, NCSTN = 0.40, NAV1 = 0.03, and SORL1 = 0.00. Although a moderate degree of spatial correlation exists between related proteins, the empirical landscape shown in Panel (b) of Figure 4 is highly rugged, underscoring the difficulty of search and discovery in pharmaceutical innovation.

FIRM HETEROGENEITY:

We use firm-level publication histories to measure heterogeneity in how organizations generalize from failure. The objective is to distinguish firms with direct, target-specific knowledge from

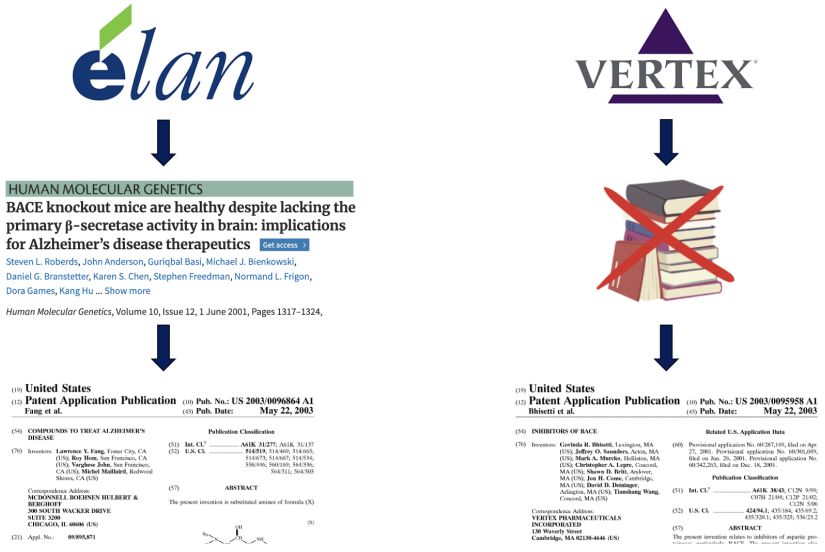
those that rely more heavily on inference across related targets. We operationalize two dimensions of the knowledge base at the firm–protein level: a target-specific *level* of experience and a cross-target *concentration* of experience. Publication data come from PubTator3, which provides machine-annotated protein mentions for PubMed records and maps them to Gene IDs and MeSH terms. We link these records to firms using author affiliation strings from Dimensions metadata and harmonize organization name variants to their parent entities. Any measurement error in this linkage would bias the results against finding systematic heterogeneity, making our estimates conservative.

More specifically, we capture prior experience for firm f, protein p, and year t by defining the following indicator:

$$\text{PriorPub}_{f,p,t} = \mathbb{I}\{\exists \text{ a PubTator3 publication by } f \text{ that studies } p \text{ with publication year } \leq t-1\}$$

This equals 1 if f has at least one prior publication on p before year t and 0 otherwise. In split-sample analyses, "higher experience" firms have $\text{PriorPub}_{f,p,t} = 1$ for the protein featured in their patents; "lower experience" firms have $\text{PriorPub}_{f,p,t} = 0$. Figure B4 shows an example from our data on two patents published on the same date, May 22, 2003. Élan Corporation has a BACE1-related article published from 2001, so $\text{PriorPub}_{\text{Élan, BACE1, 2003}} = 1$. Instead, Vertex Pharmaceuticals' first BACE1-related publication appeared in 2019, so $\text{PriorPub}_{\text{Vertex, BACE1, 2003}} = 0$. Both firms patented on BACE1 in 2003, but they differ in target-specific prior experience according to our PubTator3-based indicator.

Figure B4: Example of Patenting Firms with Differing Levels of Expertise on BACE1.



Note: The figure shows the two patents published the same day, May 22, 2003, both leveraging BACE1 as a drug target for Alzheimer's. Élan Corporation has prior publications on the protein, while Vertex Pharmaceuticals does not, thus implying differing levels of experience with the protein. See text for details.

To capture the concentration of firm experience, we follow the approach by Arts et al. (2025) and build a measure based on firms' publication portfolios. Let $\text{pubs}_{f,p,t}$ be the count of PubTator3-

indexed publications by firm f that mention protein p up to year t. We then define the share:

$$s_{f,p,t} = \frac{\text{pubs}_{f,g,t}}{\sum_{g'} \text{pubs}_{f,g',t}}, \quad \text{with} \sum_{g} s_{f,g,t} = 1,$$

and the corresponding Herfindahl–Hirschman Index (HHI): $\text{HHI}_{f,t} = \sum_g \left(s_{fg,t}\right)^2$. A higher $\text{HHI}_{f,t}$ indicates a more concentrated portfolio of protein-specific publications for firm f. According to our computational model, narrower portfolios imply a greater potential role for generalization when direct experience is limited. We operationalize this in heterogeneity tests by classifying firms with above-median $\text{HHI}_{f,t}$ as having a narrow base of experience, and firms with below-median values as having a broader one. Results are robust to alternative percentile cutoffs.

As an illustration of these measures, consider two mid-size biotechnology companies, each with the same number of patent applications in our data. The first is Anadys Pharmaceuticals, later acquired by Roche for $230 million.[11] Anadys was a leader in treatment options for hepatitis C, with a highly focused portfolio of R&D activities. Its eight peer-reviewed publications referenced only nine proteins, resulting in a high concentration of expertise: $\text{HHI}_{\text{Anadys}} = 0.1426$. In contrast, ACEA Biosciences, a similarly sized firm acquired by Agilent Technologies for $250 million,[12] specialized in developing platform tools for cell analysis rather than pursuing specific therapeutic targets. ACEA's 18 publications covered 157 proteins, yielding a much lower concentration value: $\text{HHI}_{\text{ACEA}} = 0.0103$.[13] In our empirical analysis, Anadys is classified as having a concentrated base of experience, whereas ACEA is classified as having a broad one.

[11] https://www.reuters.com/roche-to-buy-anadys-pharmaceuticals-for-230-million.html

[12] https://www.agilent.com/newsroom/presrel/2018.html

[13] A similar measure of breadth of expertise based on patent portfolios yields consistent results: Anadys' patents mention 327 proteins, whereas ACEA's mention 1,755 proteins. These unreported results confirm the robustness of the publication-based measure.

C. ADDITIONAL TABLES AND FIGURES

Table C1: Direct Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting by Clinical Trial Stage.

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Phase II and III)	0.980***	1.110***		
	(0.123)	(0.000650)		
... × Failure		-0.459**		
		(0.164)		
Post × Clinical Trial (Only Phase II)			1.202***	1.387***
			(0.105)	(0.138)
... × Failure				-0.646***
				(0.188)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	147,945,970	147,945,970

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs, while columns (3) and (4) exclude protein-disease pairs that received a Phase III clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Phase II and III)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a focal protein-disease pair; *Post × Clinical Trial (Only Phase II)*: 0/1 = 1 in all years after conclusion of the first phase 2 clinical trial targeting a focal protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C2: Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting (Alternative Samples).

	Firm Patents Targeting a Protein-Disease Pair			
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Spillover)	0.0269***	0.0302***	-0.00482***	-0.00220*
	(0.000630)	(0.000713)	(0.00105)	(0.00110)
... × Failure		-0.0146***		-0.0116***
		(0.00147)		(0.00147)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	147,979,011	147,979,011	43,281,430	43,281,430

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes all protein-disease pairs that did not receive a clinical trial, while columns (3) and (4) include only protein-disease pairs genetically related to pairs receiving a clinical trial. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C3: Spillover Effects of Clinical Trial Failures on Pharmaceutical Firm Patenting (Enrollment Size of Clinical Trial).

	Firm Patents Targeting a Protein-Disease Pair			
Sample:	Low Patient Enrollment		High Patient Enrollment	
	(1)	(2)	(3)	(4)
Post × Clinical Trial (Spillover)	0.0275***	0.0304***	0.0325***	0.0362***
	(0.000883)	(0.00107)	(0.000892)	(0.000954)
... × Failure		-0.00834***		-0.0353***
		(0.00185)		(0.00260)
Protein-disease FE	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	126,334,781	126,334,781	126,240,845	126,240,845

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. The sample in columns (1) and (2) includes only clinical trials with a below median number of patients enrolled (i.e., fewer than 50), while columns (3) and (4) include clinical trials with an above median number of patients enrolled (i.e., fewer than 50). *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C4: Spillover Effects of Clinical Trial Failures by Underlying Merit (Split Sample Regressions).

	Firm Patents Targeting a Protein-Disease Pair							
Sample:	High Merit Protein-Disease Pairs				Low Merit Protein-Disease Pairs			
Protein-Protein Distance:	1 Quartile (1)	2 Quartile (2)	3 Quartile (3)	4 Quartile (4)	1 Quartile (5)	2 Quartile (6)	3 Quartile (7)	4 Quartile (8)
Post × Clinical Trial (Spillover)	0.0180	-0.0240***	-0.0204**	-0.00771	0.00221*	-0.00257***	-0.00214***	-0.00366***
	(0.0101)	(0.00672)	(0.00691)	(0.00556)	(0.00108)	(0.000703)	(0.000609)	(0.000521)
... × Failure	-0.0433***	-0.0208*	-0.00739	-0.0131	-0.00737***	-0.00330**	-0.000822	0.000650
	(0.0144)	(0.00916)	(0.00964)	(0.00852)	(0.00199)	(0.00100)	(0.00103)	(0.000769)
Protein-disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,714,370	1,265,780	1,077,699	941,659	9,535,511	9,540,299	9,659,334	9,546,778

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level corresponding to those reported in Figure 2. The table reports standardized beta coefficients to enable comparison across split samples. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. Columns (1)-(4) include protein-disease pairs with a positive Open Target Score, while columns (5)-(8) include protein-disease pairs with an Open Target Score equal to zero. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Table C5: Spillover Effects of Clinical Trial Failures by Level of Firm Experience (Split Sample Regressions).

	Firm Patents Targeting a Protein-Disease Pair							
Sample:	Firms with High Level of Experience				Firms with Low Level of Experience			
Protein-Protein Distance:	1 Quartile	2 Quartile	3 Quartile	4 Quartile	1 Quartile	2 Quartile	3 Quartile	4 Quartile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Clinical Trial (Spillover)	0.00557**	-0.00300**	-0.00200*	-0.00168**	0.00684***	-0.00610***	-0.00543***	-0.00510***
	(0.00203)	(0.00107)	(0.000985)	(0.000634)	(0.00192)	(0.00117)	(0.00103)	(0.000949)
... × Failure	-0.00540	-0.000393	0.000684	0.000527	-0.0211***	-0.0115***	-0.00535***	-0.00255*
	(0.00314)	(0.00134)	(0.00143)	(0.000942)	(0.00230)	(0.00146)	(0.00128)	(0.00120)
Protein-disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	11,249,881	10,806,079	10,737,033	10,488,437	11,249,881	10,806,079	10,737,033	10,488,437

Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level corresponding to those reported in Figure 2. The table reports standardized beta coefficients to enable comparison across split samples. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. Columns (1)-(4) include patent applications firms with a high level of experience, while columns (5)-(8) include patent applications firms with a low level of experience. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post × Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

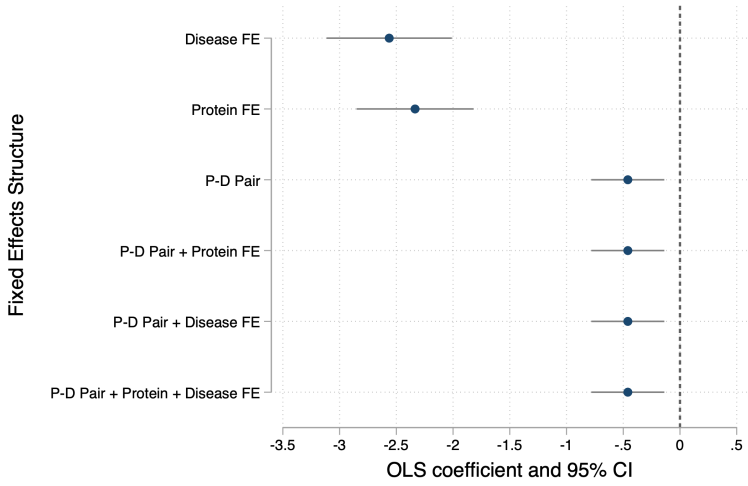
Table C6: Spillover Effects of Clinical Trial Failures by Concentration of Firm Experience (Split Sample Regressions).

	Firm Patents Targeting a Protein-Disease Pair							
Sample:	Firms with Low Concentration of Experience				Firms with High Concentration of Experience			
Protein-Protein Distance:	1 Quartile (1)	2 Quartile (2)	3 Quartile (3)	4 Quartile (4)	1 Quartile (5)	2 Quartile (6)	3 Quartile (7)	4 Quartile (8)
Post × Clinical Trial (Spillover)	0.01014***	-0.00371***	-0.00324***	-0.00357***	0.00452***	-0.00867***	-0.00626***	-0.00363***
	(0.00199)	(0.00110)	(0.000992)	(0.000767)	(0.00162)	(0.00108)	(0.000853)	(0.000706)
... × Failure	-0.0123***	-0.00473***	-0.00138	-0.000333	-0.0191***	-0.0106***	-0.00550***	-0.00319***
	(0.00282)	(0.00139)	(0.00137)	(0.00106)	(0.00189)	(0.00116)	(0.00102)	(0.000869)
Protein-disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Protein FE	YES	YES	YES	YES	YES	YES	YES	YES
Disease FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	11,249,881	10,806,079	10,737,033	10,488,437	11,249,881	10,806,079	10,737,033	10,488,437

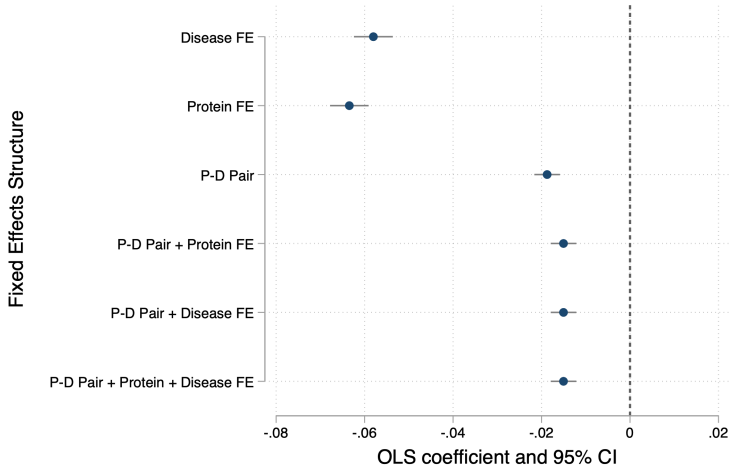
Note: *, **,*** denote significance at the 5%, 1%, and 0.1% level, respectively. Difference-in-differences panel regressions at the protein-disease-year level corresponding to those reported in Figure 2. The table reports standardized beta coefficients to enable comparison across split samples. Std. err. clustered at the protein-disease level. All models include protein-disease pair, disease, protein, and year fixed effects. Columns (1)-(4) include patent applications firms with a low concentration of publication experience, while columns (5)-(8) include patent applications firms with a high concentration of publication experience. *Firm Patents Targeting a Protein-Disease Pair*: yearly count of USPTO patent applications granted to pharmaceutical firms for inventions targeting a specific protein-disease pair; *Post* × *Clinical Trial (Spillover)*: 0/1 = 1 in all years after conclusion of the first phase 2 or 3 clinical trial targeting a genetically related protein-disease pair; *Failure*: 0/1 = 1 for clinical trials that are terminated before their natural completion. See text for details.

Figure C1: Direct and Spillover Effects from Clinical Trial Failures (Robustness to Alternative Fixed Effect Structures).

(a) *Direct Effects of Clinical Trial Failures.*



(b) *Spillover Effects of Clinical Trial Failures.*



Note: The figure presents the robustness of our main results reported in Table 2 to alternative structures of fixed effects. Each coefficient presents the OLS estimate of the interaction term with increasingly stringent specifications of gene, disease, and gene-disease pair fixed effects. Panel (a) shows the decrease in patenting for protein-disease pairs subject to failed clinical trials, depending on the fixed effect structure. The coefficients are estimated using variations of Equation 4, with the last coefficient corresponding to Column (2) of Table 2. Panel (b) shows the decrease in patenting for protein-disease pairs biologically related to the genetic targets of failed clinical trials, depending on the fixed effect structure. The coefficients are estimated using variations of Equation 5, with the last coefficient corresponding to Column (4) of Table 2. See text for details.