Multi-Project Collaborations

Charles Angelucci and Roi Orzach*

September 8, 2025

Abstract

We analyze collaborative experimentation across multiple independent domains. Each domain contains infinitely many potential projects with asymmetric benefits. In each period and domain, two players can idle, jointly explore a new project, or jointly exploit a known one, with voluntary transfers. For intermediate discount factors, treating domains as independent during experimentation is suboptimal. The optimal experimentation policy exhibits common features of collaborative experimentation: lengthy exploration, temporary project exploitation, recall of past projects, and inefficient initial or terminal idling within certain domains. We connect these findings to research on buyer-supplier dynamics and persistent productivity differences.

Keywords: Strategic Experimentation, Multi-Domain Experimentation, Relational Contracting

JEL Classification Numbers: D21, D70, D83, L25

*Angelucci: MIT Sloan School of Management, 50 Memorial Drive, Cambridge MA 02142 (e-mail: cangeluc@mit.edu); Orzach: Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215 (e-mail: orzach@bu.edu).

We thank Daniel Barron, Alexander Bloedel, Alessandro Bonatti, Renee Bowen, David Byrne, Yifan Dai, Joyee Deb, Wouter Dessein, Glenn Ellison, Dana Foarta, Guillaume Fréchette, George Giorgiadis, Robert Gibbons, Marina Halac, Johannes Horner, Matthias Lang, Alessandro Lizzeri, Jin Li, Rocco Macchiavello, Chiara Margaria, Simone Meraglia, Stephen Morris, Aroon Narayanan, Juan Ortner, Jacopo Perego, Hazhir Rahmandad, Daniel Rappaport, Patrick Rey, Kareen Rozen, Klaus M. Schmidt, Steven Tadelis, Michael Ting, Jean Tirole, Marta Troya Martinez, Jonathan Weinstein, Michael Whinston, and seminar audiences at the Oliver E. Williamson Seminar at Berkeley Haas, LMU Munich, the MIT Organizational Economics and Theory groups, the 2023 Montreal IO Conference, the Fall 2023 NBER Organizational Economics Workshop, the Organisational Economics and Leadership Workshop at Exeter University, the Paris School of Economics, the 2023 SIOE Conference, and the TSE Microeconomic Theory Workshop.

1 Introduction

In many settings, actors collaborate to experiment simultaneously in multiple domains. In buyer-supplier relationships, companies co-innovate in various product lines or geographies. In the pharmaceutical sector, an R&D alliance may combine resources across research areas. Inside firms, continuous improvement methods involve managers and workers collaborating to identify and implement improvements throughout the production process.

The success of these collaborations relies on keeping interests aligned, so that each party finds ongoing value in maintaining the partnership. In multi-domain collaborations, the ongoing value of continued participation is determined by the aggregate value across all domains of cooperation. This aggregate value–representing what parties stand to lose if cooperation ends–creates interdependencies across domains. For instance, a breakthrough in one domain will increase the parties' perceived value of the collaboration, mitigating opportunism in the other domains. As a result, parties must approach their joint experimentation in each domain of cooperation by balancing the domain-specific outcomes with the broader implications for the overall collaboration. This raises critical questions: How does multi-domain experimentation shape exploration and exploitation choices within each domain? How does the number of active domains evolve over time? Does starting with fewer domains foster cooperation? When does experimentation expand to all domains?

To address these questions, we develop a model of multi-domain collaborative experimentation and use it to interpret key findings from the applied literature that studies settings such as those mentioned above. In our model, the number of domains is exogenous, and domains are technologically independent. Each period, in each domain, two players can choose to idle, exploit a known project, or explore a new one from an infinite set of potential projects. Cooperation on a project requires the participation of both players; working individually is not an option. Project benefits are time-invariant but initially uncertain, and they may be asymmetric across players. The benefits of a project are revealed in the first period of cooperation on that project. Moreover, all projects entail a constant fixed cost for the players, during both exploration and exploitation phases. As a result, players might be reluctant to collaborate in exploring projects if they expect that their individual benefit will not exceed this cost, and they may similarly be reluctant to collaborate in exploiting a project if their realized individual benefit falls below the cost. To align incentives, players can transfer money to each other. However, these transfers are voluntary, so any experimentation policy–a rule determining for each domain whether to idle, exploit a known project, or explore a new one–must be self-enforcing.

We focus on Subgame Perfect Equilibria (relational contracts) that maximize the players' discounted cumulative joint payoffs (their "surplus"). As a first benchmark, Proposition 1 examines the single-player scenario, providing a straightforward solution in which each domain is treated independently: within each domain, exploration continues until a project's value exceeds a time-invariant threshold, after which permanent exploitation of this project is optimal. We refer to this optimal policy as the "first-best experimentation policy." This policy would also be optimal for two players if all projects benefited them equally.

However, because experimentation in our setting requires both players' participation, asymmetric project benefits create incentive challenges. We focus on the case where each project benefits only one player, with the beneficiary revealed upon first cooperation and drawn i.i.d. across projects. These asymmetric benefits create the key friction that may impede first-best experimentation, as implementing this policy requires credible promises of transfers between players. Such promises may lack credibility when players making transfers have insufficient continuation value in the collaboration. As mentioned above, since a player's continuation value equals the sum of continuation values across all domains of cooperation, experimentation choices in one domain affect all others.

In the spirit of Levin (2002, 2003), we show in Proposition 2 (i) that any optimal experimentation policy is governed solely by the value of the most profitable projects identified in each domain to date, and (ii) that a single implementability constraint, dependent on only these values and the experimentation policy, fully captures all deviation temptations across players, domains, and transfers. These results imply that any experimentation policy satisfying this constraint can be implemented through a relational contract with appropriately designed transfers. As a result, the optimal experimentation policy is characterized by a multi-dimensional Bellman equation subject to the implementability constraint. However, unlike the single-player benchmark, this constraint precludes an index characterization of the optimal policy. Despite this challenge, we establish key properties of the players' optimal experimentation policy.

Since the first-best experimentation policy treats each domain independently, we can explicitly determine the conditions under which this policy is implementable through a relational contract. Proposition 3 provides a necessary and sufficient condition: the joint value of the most valuable projects identified in each domain must be sufficiently high to ensure that the collaboration's continuation value supports the implementation of the first-best policy. For low discount factors, this condition binds, implying that, in expectation, players transition to permanently exploiting the most valuable projects found in each domain later than if they could implement the first-best from the start (Corollary 1). In some cases, this transition never occurs, as discussed below. This condition also allows a full characterization of optimal experimentation in our second benchmark, the single-domain case: exploration continues until a project's value exceeds a fixed threshold–higher than in the single-agent case–after which permanent exploitation is optimal (Corollary 2).

Next, we analyze the second-best experimentation policy, which arises when the first-best policy is not implementable in the current period. We first examine the players' exploration and exploitation decisions, abstracting from the number of domains they engage in. Notably, the player's exploitation criterion becomes dynamic. Proposition 4 shows that, whenever the distribution of project values is sufficiently rich (Definition 1), the second-best experimentation policy differs sharply from the first-best–under which explored projects are either exploited permanently or never used. Instead, with strictly positive probability, players (i) exploit projects temporarily or (ii) exploit previously abandoned projects.

We then examine the dynamics of the players' scope of experimentation–the number of domains involving either exploration or exploitation in a given period–where m represents the exogenous maximum number of potential domains. We analyze both initial and terminal (asymptotic) scope of experimentation. Starting with limited scope–such as one domain instead of m–reduces players' initial deviation temptation by a factor of m. The potential for later scope expansions, if the continuation value increases, further mitigates initial deviation temptations. However, the continuation value increases only through exploration, and exploring one domain (versus m) reduces these increases by a factor on the order of m. Proposition 5 shows that, although these opposing forces cannot generally be ranked, for large m, an initially limited scope always allows implementation over a wider range of discount factors than immediate exploration in all domains. Moreover, the continuation value of the collaboration need not increase monotonically over time: a domain's continuation value decreases when players switch from exploration to exploitation. Thus, exploiting projects in some domains may create inefficiencies in others, including being permanently idle. Proposition 6 shows that initially limited experimentation policies may never reach maximal–and thus efficient–scope asymptotically, and even policies starting with maximal scope may become permanently limited.

Further, Section 5 offers a complete characterization of the optimal experimentation policy when there are exactly two domains and each project's payoff is binary (Proposition 7). Even in this setting, the resulting optimal policy produces many of the dynamics described above. We also examine how the potential maximum scope of experimentation m impacts the feasibility and profitability of the collaboration, drawing connections to the seminal work of Bernheim and Whinston (1990) on multilateral interactions. Further, we discuss extensions of the model included in the Online Appendix.

Finally, Section 6 connects our findings to the applied literature on buyer-supplier relationships and persistent productivity differences across firms. The buyer-supplier relationships literature stresses credible joint experimentation and shows that success unfolds gradually and path-dependently. We also argue that our framework provides insights into how managerial practices can generate productivity differences among seemingly similar firms.

The rest of the paper is structured as follows. Section 1.1 reviews the relevant theoretical literature. Section 2 presents the model. Section 3 characterizes the first-best experimentation policy. Section 4 provides the main analysis, while Section 5 extends it–fully characterizing the two-domain binary-value case as well as discussing various extensions. Section 6 examines the applied literature in light of our theoretical findings. Section 7 concludes.

1.1 Related Theoretical Literature

In this section, we review the theoretical literature related to our work. We postpone the discussion of the applied literature to Section 6.

Firstly, our research connects to the literature on multi-armed bandit problems (Robbins, 1952) and on optimal search (Lippman and McCall, 1976; Weitzman, 1979), contributing to the strand with strategic interactions.[1] Bolton and Harris (1999) and Keller et al. (2005) study free-riding in experimentation (see also Hörner et al., 2022). Further, in Liu and Wong (2023) players compete to explore alternatives. In Strulovici (2010), players vote between a safe and a risky arm, with its asymmetric benefits revealed over time through experimentation (see also Anesi and Bowen, 2021). Albrecht et al. (2010) examine a search problem where a committee determines which project to exploit. Chan et al. (2018) and Reshidi et al. (2025) compare group and individual decision-making, examining the effects of static vs. sequential information acquisition and voting rules. In contrast to these papers, our setting allows for voluntary transfers and requires the players to cooperate for both exploration and exploitation. Most significantly, players experiment simultaneously across multiple domains. Experimentation across multiple domains already appears in Jovanovic

[1] Unlike typical search problems where rewards come at the end from the best explored alternative, our model allows players to benefit each time they cooperate on a project. For this reason, we use the broader term “experimentation” rather than “search.” Moreover, models of strategic experimentation with bandits often limit options to a few alternatives. We assume an infinite number of i.i.d. projects to eliminate aggregate uncertainty, making the dynamics driven purely by strategic factors.

and Rob (1990). In our framework, however, exploration or exploitation in one domain does not preclude exploration or exploitation in others, and any abandonment of a domain to pursue another is driven by strategic incentives.

The analysis of multi-domain bandits is analytically challenging. As noted by Bergemann and Välimäki (2008), "it is well known that [a Gittins] index characterization is not possible when the decision maker must or can select more than a single arm at each t," due to the possibility of recalling past arms/projects.[2] In our single-agent benchmark the hurdle disappears: infinitely many i.i.d. projects eliminate recall and restore a Gittins index (Bergemann and Välimäki, 2001). Cooperative experimentation with asymmetric payoffs re-introduces the difficulty–now the policy must be a self-enforcing equilibrium, and the cross-domain incentives again rule out any Gittins representation.

Secondly, this work relates to the literature on relational contracts (see e.g., Bull, 1987; Macleod and Malcomson, 1989; Baker et al., 1994, 2002; Levin, 2002, 2003, for early contributions).[3] Halac (2014) studies a setting in which the value of the players' relationship increases exogenously with its duration, allowing for greater efficiency. In our setting, players' experimentation endogenously shapes the continuation value of their relationship, which may or may not increase monotonically over time.

A closely related paper to ours is Chassang (2010). In his model, the agent knows which arms are productive and which are not, while the principal, at the outset, does not. Without monetary incentives, incentivizing the agent to choose productive arms is accomplished by the threat of firing the agent following failures. This dynamic makes motivating exploration progressively expensive as more productive arms are identified. Should the relationship endure, it ultimately enters an "exploitation" phase and its value stops growing. In our model, symmetrically informed players experiment across multiple domains, and transferable

[2]Bergemann and Välimäki (2008) also note that even if such an index existed, "it is normally impossible to obtain analytical solutions for the problem."

[3]Also at the intersection of the bandit and the relational contracting literatures, Urgun (2021) examines a scenario where a principal interacts with multiple agents whose publicly-observable types depend on the contracting history.

utility–apt for modeling firms–removes the need for inefficient on-path punishments.[4]

Finally, we contribute to the literature on gradualism. Watson (1999, 2002) generate gradualism through screening: each player's type (trustworthy vs. opportunistic) is private information, so partners begin with low-stakes cooperation and expand only as beliefs improve. This is not an experimentation setting–there is no exploration–exploitation margin, and the payoffs from any given interaction are known. A different benchmark is provided by symmetric-learning models in which both parties start uninformed about the match-specific surplus (e.g., Macchiavello and Morjaria, 2015). Such models can replicate expansions and contractions in scope, but many of these dynamics could also be obtained in single-agent settings with aggregate uncertainty, thereby downplaying the strategic interactions highlighted in Section 6.1. By contrast, our framework contains no private types or aggregate uncertainty and centers on experimentation: dynamics arise from incentive constraints–driven by asymmetric project benefits–that govern how players resolve exploration–exploitation trade-offs across multiple domains. This difference in mechanism yields sharply different predictions. First, the scope of cooperation in our setting need not increase monotonically–it can start large or even contrast–whereas in Watson's model scope can only grow or collapse. Second, higher patience (δ) weakens the rationale for "starting small" in our environment, whereas in Watson (1999, 2002) and in the broader dynamic screening literature (e.g., Ely and Välimäki, 2003; Acharya and Ortner, 2022) a higher δ reinforces gradualism.

In Online Appendix Section 3, we revisit the theoretical literature in light of the empirical and qualitative findings on buyer–supplier relationships and persistent productivity differences presented in Section 6, focusing on how our framework–relative to alternative theoretical models–generates and rationalizes these patterns.

[4]Introducing money in Chassang (2010), where information asymmetry plays a central role, would make the value of the relationship constant on path. We further discuss Chassang (2010) in Footnote 17 and Section 6.2. For a setting similar to Chassang (2010) but with imperfect transfers and uncertainty about the value of the relationship, see Venables (2013). For work on experimentation in principal-agent settings with commitment, see Halac et al. (2016) and Ide (2024).

2 The Setup

Two players, with a discount factor $\delta < 1$ and zero per-period outside options, have the opportunity to interact over multiple time periods $t = 1, 2, \ldots$. Their interaction spans m exogenously fixed domains–such as distinct geographical markets or product categories in a buyer-supplier relationship–where each domain j contains a countably infinite set of projects $\mathcal{P}_j$. The union of all these sets forms the total set of projects, denoted as $\mathcal{P} = \cup_j \mathcal{P}_j$, where each project within $\mathcal{P}$ is indexed by p. In each period t, and for each domain j, each player $i = 1, 2$ chooses up to one project from the set $\mathcal{P}_j$. The finite set of projects chosen by player i in period t is denoted by P_i^t. The players cooperate on the set of projects $\mathbf{P}^t = P_1^t \cap P_2^t$, following a unanimity rule, and cannot work individually on projects not included in $\mathbf{P}^t$, as both players possess indispensable and complementary assets or skills. The cardinality of this set, $|\mathbf{P}^t| \leq m$, is referred to as the scope of the players' experimentation in period t.

Each project in $\mathbf{P}^t$ costs $c > 0$ for each player and has initially unknown time-invariant value $v_p \in V \subseteq \mathcal{R}$, which is publicly observed after the first cooperation. We assume that for each project, a single player receives the entire value v_p of the project. Each project's beneficiary is fixed but initially unknown; we denote it by $x_p \in \{1, 2\}$. Both v_p and x_p are each i.i.d. across projects and domains, making all domains ex ante identical. We denote by $\alpha \in \left[\frac{1}{2}, 1\right]$ the probability that $x_p = 1$, implying that player 2 receives v_p with probability $1 - \alpha$.[5] We assume $\mathbb{E}(v_p) \geq 2c$, to ensure the first-best experimentation policy is non-empty.[6] To rule out trivial cases, we assume that the distribution of v_p is non-degenerate.

We say that a project is being "explored" when cooperated on for the first time and "exploited" when cooperated on in both the current period and at least one prior period.

[5]We could allow for more general benefit distributions, for instance with each project benefiting both players. Our main insights would still go through as long as at least some projects leave one player with a benefit below their cooperation cost c; if every project delivered benefits that exceed c for everyone, the setting would reduce to the single-agent benchmark in Section 3. Introducing this additional generality, however, would make the proofs more intricate: the optimal experimentation policy would then have to track the full set of non-Pareto-dominated projects discovered in each domain, rather than just the single project with the highest v_p as in our analysis.

[6]Without this assumption, no experimentation is optimal in the first-best for low discount factors, unnecessarily complicating our analysis of the second-best policy where the discount factor is key.

There are no intertemporal restrictions on project availability.

Further, the players exchange money twice during each period. At the beginning of each period t, the players make discretionary transfers to each other, where $w^t_{i,-i} \in \mathcal{R}^+$ denotes such a transfer from player i to player $-i$. At the end of each period t, players again make discretionary transfers to each other, where $b^t_{i,-i} \in \mathcal{R}^+$ denotes such a transfer from player i to player $-i$.[7] Finally, player i's period t payoff is equal to:

$$\pi^t_i = w^t_{-i,i} - w^t_{i,-i} + b^t_{-i,i} - b^t_{i,-i} + \sum_{p \in \mathbf{P}^t} \left(v_p \mathbb{1}_{x_p = i} - c \right), \text{ where } i \in \{1, 2\}, \tag{1}$$

and where $\mathbb{1}_{x_p=i} = 1$ if $x_p = i$ and otherwise is equal to zero.

We conclude the model's description by stating the timing of the stage game. Both players simultaneously choose their discretionary transfers $w^t_{i,-i}$. Next, both players simultaneously make their project choices P^t_i. For each project $p \in \mathbf{P}^t$, the players incur c and observe its beneficiary x_p and its value v_p, and player x_p pockets v_p. Finally, both players simultaneously choose their discretionary transfers $b^t_{i,-i}$.

Relational Contracts. A relational contract is a complete plan for the relationship. Let $h^t = (\mathbf{w}^1, \mathbf{P}^1, \mathbf{v}^1, \mathbf{x}^1, \mathbf{b}^1, \ldots, \ldots, \mathbf{w}^{t-1}, \mathbf{P}^{t-1}, \mathbf{v}^{t-1}, \mathbf{x}^{t-1}, \mathbf{b}^{t-1})$ denote the history up to date t and $\mathcal{H}^t$ the set of possible date t histories, where boldface lowercase letters indicate vectors. Then, for each date t and every history $h^t \in \mathcal{H}^t$, a relational contract describes: (i) the $\mathbf{w}^t$ transfers; (ii) the set of projects $\mathbf{P}^t(\mathbf{w}^t)$ as a function of $\mathbf{w}^t$; and (iii) the $\mathbf{b}^t(\mathbf{w}^t, \mathbf{P}^t, \mathbf{v}^t, \mathbf{x}^t)$ transfers as a function of $\mathbf{w}^t$, $\mathbf{P}^t$, and the realizations of $\mathbf{v}^t$ and $\mathbf{x}^t$. A relational contract is self-enforcing if it constitutes a Subgame Perfect Equilibrium of the repeated game. Within this class, we focus on equilibria that maximize joint surplus. Restricting to pure strategy equilibria is without loss of optimality (i) because mixing on transfers only hurts imple-

[7] We incorporate the option of monetary transfers both before and after the players' project choices, although removing either would not qualitatively affect our results. Without transfers at the beginning of each period, surplus might no longer be fully redistributed across the players without affecting incentives. Without transfers at the end of each period, incentives for the current period would rely on transfers from the subsequent period, complicating the proofs.

mentability (since the largest transfer that might be owed can only weakly increase, thereby tightening constraints) and (ii) mixing on projects leads to limited scope that can be replicated by being idle in some domains. In the event of a deviation in some period t, the players respond (i) by choosing $P_i^t = \emptyset$ and $b_{i,-i}^t = 0$ if these choices have not been made yet and (ii) by permanently breaking off their relationship (i.e., reverting to the worst equilibrium of the stage game from the next period onward). This punishment is without loss of optimality as it occurs out-of-equilibrium (c.f. Abreu, 1986).[8] Throughout, a relational contract is defined as "non-empty" if $\Pr(\sum_t |\mathbf{P}^t| > 0) > 0$.

3 First-Best Experimentation

We characterize the optimal experimentation policy for a benchmark where a single decision maker, "player 0," maximizes the sum of the payoffs of both players. This optimal experimentation policy is qualitatively identical to the one that would arise in a modified version of the model from Section 2, where each player always receives either v_p or a constant share $\frac{1}{2}v_p$. The proof of the following proposition closely follows Bergemann and Välimäki (2001) and is provided in the Appendix, along with proofs for all other statements omitted from the main text.

Proposition 1 (**First-Best Experimentation Policy**)
For each domain j and period t, player 0 adopts the following experimentation policy: if a previously-explored project p has the highest value and $v_p \geq v^0(\delta)$, exploit it; Otherwise, explore a new project. The threshold $v^0(\delta)$ is chosen from the closure of the support of project values, $v^0 \in \mathrm{cl}\,(V)$, and it is increasing in δ.

Player 0 treats domains identically due to additive payoffs and ex ante symmetry. The threshold v^0 reflects their choice in each domain between exploiting the best known project

[8]Alternatively, players could maintain the equilibrium but allocate all surplus to the non-deviator. This provides identical incentives and, being Pareto optimal, is less prone to renegotiation.

or exploring for a better one.[9] Exploitation is permanent, as no new information is gained and rejected projects are never revisited given the infinite supply. Finally, as the discount factor rises, exploration becomes more valuable, driving the comparative statics of v^0.

In summary, the first-best policy maximizes experimentation scope, with exploration and exploitation in each domain following a time-invariant threshold. We now analyze the model from Section 2, identifying when the fixed-threshold logic breaks down and how that shapes exploration, exploitation, and scope over time.

4 Main Analysis

In Section 4.1, we characterize the class of optimal relational contracts on which the analysis focuses and establish a necessary and sufficient condition for an experimentation policy to be implementable by an optimal relational contract. In Section 4.2, we provide the conditions under which the players can implement the first-best policy. In Section 4.3, we characterize key properties of the optimal experimentation policy when they are unable to implement the first-best policy.

4.1 Optimal Experimentation Policies: Implementability

Surplus-maximizing relational contracts depend on the players' beliefs about the projects, denoted by $\mu^t(h^t) := \{\Delta(v_p, x_p)|h^t\}_{p\in\mathcal{P}}$. We show that there exist surplus-maximizing relational contracts that condition on h^t only through $\mu^t(h^t)$. Moreover, restricting attention to relational contracts specifying the same continuation equilibrium following any two on-path histories h_1^t and $h_2^{t'}$ leading to the same beliefs μ is without loss of optimality, since the only history-dependent outcome that alters the set of continuation equilibria are the beliefs. The continuation equilibria prescribed by such surplus-maximizing relational contracts are also

[9] When the support of v_p is convex, the threshold $v^0(\delta)$ is uniquely pinned down. If the support contains gaps, several cut-offs that lie in those gaps could yield identical policies. To avoid this inessential multiplicity, we always choose $v^0(\delta)$ inside the closure of $\mathrm{Supp}(v_p)$.

surplus-maximizing; otherwise, non-surplus-maximizing continuation equilibria could be replaced with surplus-maximizing ones, with appropriate transfers to maintain incentives. We refer to such relational contracts as optimal. The following proposition formalizes this characterization and provides a necessary and sufficient condition for an experimentation policy $\hat{\mathbf{P}} : \{\Delta(v_p, x_p)\}_{p\in\mathcal{P}} \to \mathcal{P}$ to be implementable by an optimal relational contract. This condition depends on the players' continuation value, $\mathcal{C}\left(\hat{\mathbf{P}}(\cdot), \mu^t\right) = \mathbb{E}\left(\sum_{t'=t+1}^{\infty} \delta^{t'-t}\left(\pi_1^{t'} + \pi_2^{t'}\right)\right)$, evaluated at belief μ^t under policy $\hat{\mathbf{P}}(\cdot)$.

Proposition 2 (**Optimal Relational Contracts**)

- *For any surplus-maximizing relational contract, there exists an alternative surplus-equivalent relational contract such that (i) for all t and for all on-path histories $h^t \in \mathcal{H}^t$, the continuation equilibrium is surplus maximizing, and (ii) for any two on-path histories h_1^t and $h_2^{t'}$, if $\mu^t(h_1^t) = \mu^{t'}\left(h_2^{t'}\right)$, then the relational contract specifies the same continuation equilibrium following these histories.*

- *There exists an optimal relational contract that implements an experimentation policy $\hat{\mathbf{P}}(\cdot)$ if and only if the following inequality holds for all on-path $h^t \in \mathcal{H}^t$:*

$$\sum_{i=1}^{2} \sum_{p\in\hat{\mathbf{P}}(\mu^t)} \max\left(0, c - \mathbb{E}\left(v_p \mathbb{1}_{x_p=i} | \mu^t\right)\right) \leq \mathcal{C}\left(\hat{\mathbf{P}}(\cdot), \mu^t\right). \tag{2}$$

The proof of this proposition extends the work of Levin (2003). In our setting, despite the stochastic nature of the players' continuation value, we show that considering its expectation is sufficient to characterize the experimentation policies that can be implemented by a relational contract.

The intuition for the first statement was provided above the proposition. Next, recall that the main tension faced by the players is that the experimentation policy which maximizes their joint surplus involves the selection of projects that do not benefit both players. Inequality (2) states that for an optimal relational contract to implement an experimentation policy everywhere on path, the continuation value induced by this policy must exceed the total reneging temptation across players and projects in all periods and histories. In turn, the total reneging temptation is the sum across players and projects of a project's reneging temptation to a player, which is either zero if the project generates a positive net expected gain, or equal to the magnitude of the net expected loss. The sum is across projects because, for any beliefs μ, each player can deviate by selecting any subset of $\hat{\mathbf{P}}(\mu)$. This condition is necessary for the relational contract to constitute an equilibrium. In the proof, we show that the presence of money also ensures sufficiency.

The proposition implies that characterizing the optimal relational contract reduces to determining the players' optimal experimentation policy, subject to Inequality (2) holding along the equilibrium path.[10] This simplification arises because all transfers cancel out in both the joint surplus expression and the right-hand side of (2). Building on this observation, we now state the corresponding optimization problem.

The optimal experimentation policy in any period depends only on the highest-valued projects identified in each of the m domains, denoted $\hat{v}_1, \ldots, \hat{v}_m$, with $\hat{v}_j := 0$ if no projects have been explored in domain j. Players never exploit a project with lower value than another, as this would reduce joint payoff and tighten Inequality (2). Thus, tracking $\hat{\mathbf{v}} := (\hat{v}_1, \ldots, \hat{v}_m)$ is sufficient to represent players' beliefs about the projects. For each j, they choose one of three actions: remain idle ($a_j = 0$), explore a new project ($a_j = 1$), or exploit the highest-valued known project ($a_j = 2$). The experimentation policy is then determined by solving the following Bellman equation, where $B(\hat{\mathbf{v}})$ represents the players' joint surplus, and where $\mathcal{C}(\cdot)$ is redefined in terms of $(\mathbf{a}, \hat{\mathbf{v}})$ instead of $(\hat{\mathbf{P}}(\cdot), \mu^t)$:

[10] The proof constructs one feasible schedule of transfers that implements the experimentation policy, but this schedule is generally not unique: many alternative transfer vectors can implement the same project choices. Hence one must think in terms of aggregate reneging temptations, rather than those of a specific player.

$$B(\hat{\mathbf{v}}) = \max_{\mathbf{a}\in\{0,1,2\}^m} \left\{ \sum_{j=1}^{m} \left[\mathbb{1}_{a_j=1}\mathbb{E}(v_p - 2c) + \mathbb{1}_{a_j=2}(\hat{v}_j - 2c) \right] + \mathcal{C}(\mathbf{a}, \hat{\mathbf{v}}) \right\} \tag{3}$$

$$\text{subject to:} \quad \sum_{j=1}^{m} \left[\mathbb{1}_{a_j=1} \max\{0, c - (1-\alpha)\mathbb{E}(v_p)\} + \mathbb{1}_{a_j=2} c \right] \leq \mathcal{C}(\mathbf{a}, \hat{\mathbf{v}}). \tag{4}$$

Notably, Inequality (4) aggregates incentives across domains, creating interdependencies that are central to our analysis. These interdependencies complicate the characterization of optimal experimentation, as we explain below. Also, we caution against the intuition that the continuation value necessarily grows over time. While both the players' joint surplus, $B(\cdot)$, and the continuation value under a fixed policy, $\mathcal{C}(\mathbf{a}, \cdot)$, increase, the equilibrium continuation value, which with an abuse of notation we denote by $\mathcal{C}(\hat{\mathbf{v}}) := \mathcal{C}(\arg\max_{\mathbf{a}} B(\hat{\mathbf{v}}), \hat{\mathbf{v}})$, need not be monotonic. This non-monotonicity arises even under the first-best policy of Proposition 1. To see why, let $B^0(\cdot)$ and $\mathcal{C}^0(\cdot)$ denote the equilibrium Bellman equation and continuation value, respectively, for first-best experimentation (which, recall, treats domains independently). For $\hat{v} < v^0$, we have $B^0(\hat{v}) = \mathbb{E}(v_p) - 2c + \mathcal{C}^0(\hat{v})$, while for $\hat{v} > v^0$, $B^0(\hat{v}) = \hat{v} - 2c + \mathcal{C}^0(\hat{v})$. Since $v^0 > \mathbb{E}(v_p)$ (otherwise, permanently exploitation would not be optimal), it follows that $\mathcal{C}^0(\cdot)$ drops discontinuously when switching from exploration to exploitation. This decline in continuation value, which reflects reduced future experimentation, further complicates the analysis, since Inequality (4) does not necessarily become easier to satisfy over time.

4.1.1 Challenges in Characterizing Optimal Experimentation

As discussed in Section 1.1, our setting does not admit a Gittins Index characterization. More generally, any characterization of the optimal experimentation policy is generally infeasible. First, the choice set is discrete, which precludes the use of continuous optimization methods. Second, due to Inequality (4), this multi-dimensional optimization problem cannot be decomposed into m independent optimization problems. As a result, the curse of dimensionality arises for $m > 1$ due to two interrelated reasons. First, even for $m = 2$, the choice set in any given period t consists of 9 options (or 5, under symmetry), and this number grows exponentially with m. Second, determining whether a given choice is feasible and optimal requires knowledge of $B(\hat{\mathbf{v}}')$ for all $\hat{\mathbf{v}}' \geq \hat{\mathbf{v}}$, and subsequently, computing its respective integral over all possible future values of $\hat{\mathbf{v}}'$ for each choice $\mathbf{a}$ to evaluate $\mathcal{C}(\mathbf{a}, \hat{\mathbf{v}})$. If the support of v_p is finite, the problem could, in principle, be solved using "backward induction" on the Bellman equation. However, this approach is analytically feasible only when both m and the cardinality of the support of v_p are very small. Accordingly, and in the spirit of Chassang (2010), Section 5.1 provides a full characterization of the two domain, binary-project-value case (which in fact Chassang (2010) could not achieve, due to difficulties we explain in Footnote 23).

4.2 Implementability of First-Best Experimentation

We provide necessary and sufficient conditions on the values $\hat{v}_1, \ldots, \hat{v}_m$ under which the players can implement the first-best experimentation policy described in Proposition 1 in the current and in all subsequent periods. We refer to this outcome as "implementing the first-best experimentation policy." As we will show, there may exist a period $t' > t$ such that the players can implement the first best in period t' and all subsequent periods, but not in the earlier period t.

Inequality (2) implies that there exists a threshold $\tilde{v}$, equal to $c\,(1+\delta)\,/\delta$, which corresponds to the minimum project value required for a project's exploitation to be sustainable in equilibrium when there is only one domain of cooperation ($m = 1$). Using this threshold $\tilde{v}$, we now provide the conditions on $\hat{v}_1, \ldots, \hat{v}_m$ under which the players can implement the first-best experimentation policy.

Proposition 3 (**Nec. and Suff. Condition for First-Best Experimentation**)
In any optimal relational contract and for any period t, the players implement the first-best experimentation policy for all $t' \geq t$ if and only if:

$$h(\hat{v}_1, \ldots, \hat{v}_m) := \frac{1}{m} \sum_{j=1}^{m} \max\{\hat{v}_j, v^0\} \geq \tilde{v} := c\frac{1+\delta}{\delta}. \tag{5}$$

As a result, there exists a threshold $\delta^0 < 1$ such that the players implement the first-best experimentation policy from period 1 onward if and only if $\delta \geq \delta^0$.

When Inequality (5) holds, the continuation value of the relationship is high enough to support the first-best policy. Since players can pool relational incentives across domains, the condition requires that the *average* across domains of the maximum between each domain's best-known project and the threshold v^0 exceeds $\tilde{v}$. The function $h(\hat{v}_1, \ldots, \hat{v}_m)$ differs from the arithmetic mean of $\hat{v}_1, \ldots, \hat{v}_m$ because (i) under the first-best policy, players explore rather than exploit projects below v^0, and (ii) exploration adds to continuation value. Moreover, the condition $v^0 \geq \tilde{v}$ is necessary and sufficient for Inequality (5) to hold from period 1 onward. Finally, the function $v^0(\delta) - \tilde{v}(\delta)$ satisfies a single-crossing property in δ, implying the existence of a threshold δ^0.[11]

Proposition 3 allows us to give necessary and sufficient conditions under which the players cease all exploration and transition to exploiting the most valuable project discovered in each domain, provided that they are already implementing the first-best experimentation policy. We refer to this outcome as "permanent exploitation."

Corollary 1 (**Nec. and Suff. Condition for Permanent Exploitation**)
In any optimal relational contract, the players permanently exploit projects with values $\hat{v}_1, \ldots, \hat{v}_m$ if and only if $\hat{v}_j \geq v^0$ for all j and the average of $\hat{v}_1, \ldots, \hat{v}_m$ exceeds $\tilde{v}$.

Proof of Corollary 1. Proposition 3 establishes that these conditions are jointly sufficient. Fixing $\hat{\mathbf{v}}$, the continuation value associated with permanent exploitation of $\hat{\mathbf{v}}$ is weakly lower

[11]Proposition 1 establishes that the threshold, $v^0(\delta)$, monotonically increases in δ, while the definition of $\tilde{v}$ implies that $\tilde{v}(\delta)$ monotonically decreases in δ.

than that under the first-best policy at $\hat{\mathbf{v}}$. Hence, if the players are able to permanently exploit $\hat{\mathbf{v}}$, they can also implement the first-best policy. This implies that these conditions are not only sufficient but also jointly necessary. □

The conditions stated in Corollary 1 imply that, in expectation, the players achieve the permanent exploitation outcome weakly later than if they could follow the first-best experimentation policy from period 1 onward. This delay relative to the first-best is strictly positive when $\delta < \delta^0$. In fact, as we will show in Proposition 6, permanent exploitation in all domains of cooperation is not guaranteed to occur.

We conclude by noting that the conditions listed in Corollary 1 fully characterize the players' optimal experimentation policy for our second natural benchmark: a single-domain collaboration. When there is only one domain (and the optimal relational contract is non-empty), the players face a simple decision in each period: either to exploit the best project found thus far or to explore a new project. The exploitation threshold in this setting is time-invariant, as the players' continuation value depends solely on the value of the best project in this single domain.

Corollary 2 (**Single-Domain Experimentation Benchmark**)
Fix $m = 1$. There exists a threshold $\delta^* \leq \delta^0$ such that the optimal relational contract is non-empty if and only if $\delta \geq \delta^*$. Furthermore, in any non-empty optimal relational contract, there exists a threshold $v^*(\delta) = \max\{\tilde{v}(\delta), v^0(\delta)\}$ such that the players explore projects until they find a project p with an associated value $v_p \geq v^*$. Once they find such a project, the players exploit it in all subsequent periods.

In this subsection, we have provided the conditions on the best projects found in each domain under which the players implement the first-best experimentation policy.[12] We have also shown that, if δ is not sufficiently high, the players will initially be unable to

[12] $\delta^* < \delta^0$ when $m = 1$ whenever the distribution of v_p satisfies *richness* at δ^0 (see Definition 1). Proposition 4 formally establishes that, under richness at δ^0, the inequality $\delta^* < \delta^0$ holds for any m. Richness at δ^0 is satisfied by any continuous distribution with convex support. Additionally, it can be shown that $\delta^*(\alpha)$ is increasing in α, with $\delta^*(1/2) = 0$. It then follows that $\delta^*(\alpha) < \delta^0$ for all α sufficiently close to 1/2.

implement the first-best policy. We now proceed to characterize key properties of the players' experimentation policy in the periods that precede a possible transition to the first-best policy.

4.3 Second-Best Experimentation

We now analyze the players' optimal experimentation policy when they cannot implement the first-best policy in the current period– what we term "second-best experimentation." This analysis focuses on the case where the maximal potential scope of experimentation, m, is strictly greater than 1 (for the case $m = 1$, see Corollary 2). Throughout, let δ^* denote the critical discount factor above which the players' experimentation set is nonempty (its existence and properties are established in the proof of Proposition 5).

The players' exploration and exploitation decisions within their active domains of collaboration are intertwined with their choices of which domains to engage in. To disentangle these dynamics, we analyze them separately: Section 4.3.1 focuses on exploration and exploitation, keeping scope decisions in the background, while Section 4.3.2 reverses the focus.

4.3.1 The Dynamics of Exploration-Exploitation Decisions

Under the first-best policy, each domain is treated independently and identically, with a time-invariant threshold for project exploitation. This time-invariance ensures that once a project is exploited or deemed unworthy of exploitation, the decision is permanent. For collaborative experimentation, players aggregate incentives across all domains. This observation implies that the criterion used to determine project exploitation may be dynamic. As a result, the players may exploit a project temporarily, and further, they may recall a project they previously chose not to exploit–behaviors that we examine in detail here.

These behaviors cannot arise if the distribution of project values v_p is too coarse. For example, in the binary–value case of Section 5.1, the players permanently exploit only high-value projects, which sharply limits (though does not entirely rule out) second-best dynamics.

To admit richer exploration–exploitation patterns, we therefore introduce a *richness* condition on the support of v_p, a mild requirement that—for example—is automatically met for any δ whenever v_p is continuously distributed with convex, unbounded support.

Definition 1 (**Richness**) *The distribution of project values v_p satisfies richness at discount factor δ if there exists $\epsilon > 0$ such that the distribution has strictly positive density on the interval $(\tilde{v}(\delta) - \epsilon, \tilde{v}(\delta) + \epsilon)$.*

Distributions that satisfy richness at δ satisfy the following two properties. First, because $\tilde{v}(\cdot)$ is continuous in δ, the absence of a mass point near $\tilde{v}$ prevents a continuous change in δ from discontinuously affecting the probability of finding a project suitable for exploitation.[13] This property can be used to show that for $\delta \uparrow \delta^0$, the players can engage in exploration(s) despite a (continuously) lower continuation value. Further, that the density is strictly positive within an ϵ-ball ensures that, when $m = 1$, there exist (i) projects that the players can exploit while satisfying Inequality (2) with slack and (ii) other projects the players would want to exploit but their exploitation would violate Inequality (2). The next proposition shows the implications of these two features when $m > 1$.

Proposition 4 (**Temporary Exploitation and Recall of Projects**)
If v_p satisfies richness at δ^0, then $\delta^* < \delta^0$. Further, fix $\delta \in [\delta^*, \delta^0)$ and $m > 1$. If v_p also satisfies richness at δ, then with strictly positive probability, at least one of the following holds:

1. *There exists a t where the players choose to exploit a project in period t, but later decide not to exploit the same project in some period $t' > t$.*

2. *There exists a t where the players choose not to exploit a project in period t, but later decide to exploit the same project in some period $t' > t$.*

[13] While this property ensures $B^0(\cdot)$ is continuous, it does not guarantee continuity of $B(\cdot)$. We discuss the challenges posed by these discontinuities in Section 4.3.2.

We provide intuition for why these two seemingly suboptimal behaviors are optimal by examining two specific examples with $m = 2$. The proof establishes that these behaviors necessarily occur with strictly positive probability.

The first statement is illustrated by the following scenario. Suppose $\hat{v}_1 \geq \hat{v}_2$, with both values high enough to sustain maximal scope but insufficient to implement the first-best. If $\hat{v}_1$ is large, the players exploit the project in domain 1 and explore domain 2. If this exploration uncovers a project slightly better than $\hat{v}_1$, roles reverse: domain 2 is now exploited while domain 1 is explored. Section 5.3 simulates this behavior in a parameterized example.

To see the second statement, consider a low δ that prevents exploiting projects just above v^0. Suppose the scope of experimentation is maximal (e.g., $\alpha = 1/2$).[14] If the first round of exploration yields two projects slightly above v^0, the players must continue exploring. But if a new project is sufficiently valuable, it raises the continuation value enough to support first-best experimentation—making it optimal to return to a period-1 project initially deemed unworthy of exploitation.

Temporary project exploitation or project recall are common in experimentation settings, and can arise due to various factors, including project characteristics that are non-stationary or the presence of a finite number of projects. Our analysis shows that strategic interactions *alone* can drive these behaviors, even in the absence of the aforementioned technological factors.

4.3.2 The Dynamics of the Scope of Experimentation

When the discount factor is high enough ($\delta \geq \delta^0$), the first–best policy is implementable from the outset, so the number of active domains is always maximal (the scope $|\mathbf{P}^t| = m$ $\forall t$). We therefore focus on the *second–best* region $\delta \in [\delta^*, \delta^0)$ to examine the dynamics of the players' scope of experimentation.

We say that an experimentation policy is "initially maximal" if $|\mathbf{P}^1| = m$, and "initially

[14]When $\alpha = 1/2$, exploration occurs in each domain in the static equilibrium, so any optimal experimentation policy has maximal scope throughout.

limited" otherwise. Likewise, it is "terminally maximal" if $\lim_{t\to\infty} |\mathbf{P}^t| = m$, and "terminally limited" otherwise. An initially maximal policy is always preferred when feasible, as exploring all domains yields immediate gains ($\mathbb{E}(v_p) \geq 2c$) and maximizes the continuation value. The key question is whether a limited policy can be implemented when a maximal one cannot. Intuitively, starting with limited scope and expanding later may be more sustainable because: (i) it reduces early reneging temptations while preserving continuation value, and (ii) early successes can support both exploitation and further exploration. We show this logic holds when the maximum scope m exceeds a threshold, but may fail below it.

To build intuition, we compare the period-1 version of Inequality (2) under two policies: an initially limited policy where players explore domain 1 only, and an initially maximal policy where all domains are explored:

$$c \leq \delta \int B(\hat{v}_1, 0, \ldots, 0|\delta) d\hat{v}_1, \tag{6}$$

$$m \cdot c \leq \delta \int B(\hat{v}_1, \ldots, \hat{v}_m|\delta) d\hat{v}_1, \ldots, d\hat{v}_m, \tag{7}$$

where $B(\cdot)$ was defined in Equation (3), and where we make explicit its dependence on δ, which will be central to the discussion below. We focus on period 1, abstracting from long-run dynamics. Since the right-hand side of (7) increases with δ, there exists a cutoff $\bar{\delta}(m) \in (0, 1)$ below which the constraint fails. The key question is whether (6) still holds for some $\delta < \bar{\delta}(m)$.

Assume, incorrectly, that $B(\cdot \mid \delta)$ is continuous in δ (it is not, since optimal experimentation itself is discontinuous). Under this assumption, and using (6)-(7), the initially limited policy is optimal just below $\bar{\delta}$ if and only if:

$$\int B\left(\hat{v}_1, 0, \ldots, 0|\bar{\delta}(m)\right) d\hat{v}_1 > \frac{1}{m} \int B\left(\hat{v}_1, \ldots, \hat{v}_m|\bar{\delta}(m)\right) d\hat{v}_1, \ldots, d\hat{v}_m. \tag{8}$$

The right-hand side of the inequality reflects the average per-domain surplus from period 2 onward under the initially maximal policy and is, intuitively, bounded above by the first-best surplus in a single domain. The left-hand side captures the total surplus under the initially limited policy and, by monotonicity of the Bellman equation, is bounded below by:

$$B\left(\mathbf{0}|\bar{\delta}(m)\right) = m\mathbb{E}\left(v_p - 2c\right) + \mathcal{C}\left(\mathbf{0}\right) \geq m\mathbb{E}\left(v_p - 2c\right) + m\left(c - (1-\alpha)\mathbb{E}\left(v_p\right)\right), \tag{9}$$

where these steps follow from the definition of the expected continuation value, and Inequality (2), respectively. If $c - (1-\alpha)\mathbb{E}\left(v_p\right) \leq 0$, then exploration is an equilibrium of the stage game, implying $\bar{\delta}(m) = \delta^*(m) = 0$. Thus, we assume, $c - (1-\alpha)\mathbb{E}\left(v_p\right) > 0$, implying this lower bound grows with m, Inequality (8) holds for large enough m, implying that initially limited policies are implementable at lower discount factors than maximal ones. We formalize this argument in the Appendix, accounting for the possible discontinuity of $B(\cdot \mid \delta)$ in δ.

In contrast, for small values of m, an initially limited experimentation policy may or may not be easier to implement than an initially maximal one. As shown in Proposition 4, exploration and exploitation decisions are optimally co-determined across domains. Thus, delaying exploration in domain j not only reduces surplus in that domain but may also lower surplus in other domains. In the Online Appendix, we explicitly construct an example which shows that when m is small, this advantage of initially maximal policies can outweigh the benefits of limited scope discussed above. These intuitions are formalized in the next proposition.

Proposition 5 (**Initial Scope of Experimentation**)
Fix $m > 1$. There exist thresholds $0 \leq \delta^*(m) \leq \bar{\delta}(m) \leq \delta^0$ such that any optimal relational contract is as follows:

(i) ***Initially maximal scope.*** *If $\delta \geq \bar{\delta}(m)$, the players start by exploring all domains: $|\mathbf{P}^1| = m$.*

(ii) ***Initially limited scope.*** *If $\delta \in [\delta^*(m), \bar{\delta}(m))$, the players start by exploring fewer than m domains: $|\mathbf{P}^1| < m$.*

(iii) ***No experimentation.*** *If $\delta < \delta^*(m)$, no experimentation can be supported; the scope is zero in every period.*

The threshold $\bar{\delta}(m) > 0$ if and only if $(1-\alpha)\,\mathbb{E}\,(v_p) < c$. Further, define $m^* := \sup\{m : \bar{\delta}(m) = \delta^*(m)\}$. If $\bar{\delta}(m) > 0$ and $\liminf_{m\to\infty} \Pr\left(v_p > \tilde{v}\left(\bar{\delta}(m)\right)\right) > 0$, m^* is finite. Consequently, for all $m > m^*$ the interval $[\delta^*(m), \bar{\delta}(m))$ is non-empty, so there is a range of discount factors for which an initially limited scope is strictly optimal.

If $(1-\alpha)\mathbb{E}(v_p) \geq c$, then exploration is a Nash equilibrium of the stage game. In this (trivial) case, players are active in all m domains every period. We therefore focus on the more interesting case where $(1-\alpha)\mathbb{E}(v_p) < c$, so that $\bar{\delta}(m) > 0$.

To show that an experimentation policy with initially limited scope is optimal for large m, we require a richness condition milder and implied by that in Definition 1: at the cutoff discount factor $\bar{\delta}(m)$, there must exist projects that would still be exploited if δ were slightly reduced. Formally, we require that $\Pr\left(v_p > \tilde{v}(\bar{\delta}(m))\right) > 0$ for large m, which is implied by $\liminf_{m\to\infty} \Pr\left(v_p > \tilde{v}(\bar{\delta}(m))\right) > 0$.

Without this condition, no project would be worth exploiting once $\delta < \bar{\delta}(m)$, and since $(1-\alpha)\mathbb{E}(v_p) < c$, exploration would also not be sustainable. Although the proposition only requires this richness condition to hold for sufficiently large m, it is automatically satisfied for all m whenever either (i) the support of v_p is unbounded, or (ii) α is sufficiently high.[15] Finally, while a general closed-form characterization of m^* is unavailable, in the parameterized example in Section 5.1 we show that $m^* = 2$.

The previous proposition established results on the players' initial scope of experimentation but did not address its long-term dynamics. We now present findings on their terminal scope.[16] In doing so we ask, does the terminal scope become maximal and how does the

[15]To build intuition for (ii), note that when $\alpha = 1$, Inequality (2) is identical for exploration and exploitation. Hence, if players can sustain exploration at date 1, they can also sustain exploitation of sufficiently high-value projects. By contrast, when α is lower, exploration becomes easier to sustain–in fact, it may be possible to sustain indefinitely even in the absence of feasible exploitation.

[16]While we prove Proposition 6 using backwards induction on the Bellman equation for a finite-support distribution, the results do not rely on the support of v_p being bounded.

terminal scope relate to the initial scope?

Proposition 6 (**Terminal Scope of Experimentation**)

For each of the two statements below there exists a non-empty open set of primitives (consisting of a distribution of project values v_p and values of δ, c, α, m) where the statement holds.

1. *Optimal experimentation policies are both initially limited and, with strictly interior probability, terminally limited.*

2. *Optimal experimentation policies are both initially maximal and, with strictly interior probability, terminally limited.*

For the players' scope of experimentation to be terminally limited, there must exist a set of vectors of project values reached with positive probability on path and a subset of domains $s \subset \{1, \ldots, m\}$ such that:

a) players can only permanently exploit projects in s due to insufficient continuation value;

b) exploring any domain $j \in \{1, \ldots, m\} \setminus s$ requires foregoing exploitation in one or more domains in s due to insufficient continuation value; and

c) players prefer exploiting all projects in s over delaying some exploitations to explore additional domains.

To prove the first statement of the proposition (respectively, the second statement), in the Appendix we show that a), b), and c) hold simultaneously under an initially limited (respectively, initially maximal) policy. Intuitively, these dynamics arise only when $\hat{\mathbf{v}}$ is high enough for a) and c) to hold but low enough for b) to be satisfied.[17]

[17] The fact that terminally limited scope may arise for intermediate values of $\hat{\mathbf{v}}$—and consequently for intermediate values of the relationship—is reminiscent of Chassang (2010)'s result, where exploration may cease when some but not all productive actions have been "revealed," leaving the value of the relationship in

In this subsection, we analyzed the dynamics of the players' scope under second-best experimentation. We showed that for intermediate discount factors and large maximal potential scope m, the players find it optimal to begin with limited scope, an approach made credible by the possibility of many subsequent scope expansions created by the discovery of valuable projects in the early domains of cooperation. Because the discovery of such projects is path-dependent, the players may end with a permanently limited and thus inefficient scope of experimentation.

5 Further Analysis and Extensions

This section extends our analysis in multiple directions. First, we fully solve the two-domain binary-value case. Second, we examine how the maximum potential scope of experimentation influences its feasibility and profitability. Third, we graphically illustrate some of the key dynamics of the model. Finally, we explore several simple extensions in which the domains of cooperation are not identical or independent.

5.1 The Two-Domain Binary-Value Case: Full Characterization

In this section, we provide a full characterization of the players' optimal experimentation policy by assuming (i) two domains ($m = 2$), (ii) binary project values $v_p \in \{\underline{v}, \bar{v}\}$ with $\bar{v} > \underline{v}$, and (iii) $\alpha = 1$, so that player 1 is the sole beneficiary of every project.[18]

Proposition 2 and the subsequent analysis show that we may, without loss, restrict attention to experimentation policies that depend in each period only on the current vector of

an intermediate range. Despite the differences in setting, the core intuition is similar: conducting additional exploration requires halting the exploitation of an existing project. In our setting, the newly explored project cannot be exploited in the current period due to b). In Chassang's setting, the absence of transferable utility means that exploring an additional action may require terminating the relationship, thereby sacrificing some future exploitation. The difficulty in computing the endogenous loss from these forgone exploitations in closed form is precisely what hinders analytical characterizations in both settings.

[18]Restricting attention to scenarios with high values of α allows us to focus squarely on the most interesting dynamics, because when α is low, scope is maximal throughout $\forall\delta$. Therefore, setting $\alpha = 1$ simplifies expressions without sacrificing any qualitative insights.

values $(\hat{v}_1, \hat{v}_2)$. Because the domains are ex-ante identical, two policies that differ only by a relabelling of the domains are payoff-equivalent; we therefore speak of an experimentation policy as "optimal" only up to symmetry. It is then convenient to reduce the state space to $n \in \{0, 1, 2\}$, the number of domains in which a $\bar{v}$-value project has already been found (noting that further exploration in such a domain is never optimal). An experimentation policy prescribes, for each n, how many domains to explore and how many to exploit.

Under this convention, the next proposition shows that any non-empty optimal relational contract can implement only three distinct experimentation policies. These policies are then ordered by (i) the total surplus they generate and (ii) the smallest discount factor δ required for their implementation. Table 1 provides a concise summary of these policies.

Policy	$\boldsymbol{n = 0}$	$\boldsymbol{n = 1}$	$\boldsymbol{n = 2}$
First Best (FB)	explore, explore	exploit $\bar{v}$, explore	exploit $\bar{v}$, exploit $\bar{v}$
Gradual-Exploit (GE)	explore, idle	exploit $\bar{v}$, explore	exploit $\bar{v}$, exploit $\bar{v}$
Gradual-Wait (GW)	explore, idle	idle $\bar{v}$, explore	exploit $\bar{v}$, exploit $\bar{v}$

Table 1: Optimal experimentation policies with two domains and binary values. For each n (number of domains in which a $\bar{v}$-value project has already been found), the table lists the prescribed actions in the two domains.

Proposition 7 (**Optimal experimentation: two domains and binary values**) *Fix $\alpha = 1$, $m = 2$, and $v_p \in \{\underline{v}, \bar{v}\}$ with $\underline{v} < \bar{v}$. Every non–empty optimal relational contract implements exactly* one *of the three experimentation policies listed in Table 1. They are ordered as follows:*

(i) Surplus: FB $>$ GE $>$ GW.

(ii) Implementability: each experimentation policy $e \in \{\mathrm{FB}, \mathrm{GE}, \mathrm{GW}\}$ is feasible if and only if the discount factor satisfies $\delta \geq \delta^e$. For every $\left(c, \underline{v}, \bar{v}, \Pr[v_p = \bar{v}]\right)$, $\delta^{\mathrm{FB}} > \delta^{\mathrm{GE}}$.

Further, δ^{GE} is strictly higher than δ^{GW} on one non-empty open parameter set and is strictly lower than it on another.

The proof of Proposition 7 is presented here, as its constructive structure helps with intuition. We note that, by definition, $\delta^{\mathrm{FB}} = \delta^0$, as defined in Proposition 3.

Proof. We begin with three observations:

Observation 1 *Exploiting a $\underline{v}$ project is not optimal: one more round of exploration yields weakly higher current surplus and a strictly higher continuation value.*

Observation 2 *If a $\bar{v}$ project has been identified in a domain, continued exploration in that domain is suboptimal.*

Observation 3 *When $m = 1$, the players can sustain the exploitation of a project with value $\bar{v}$–with Inequality (2) holding strictly–if and only if:*

$$c < \frac{\delta}{1-\delta}(\bar{v} - 2c). \tag{10}$$

If this condition fails, no non-empty relational contract can be sustained for $m = 1$ or $m = 2$, since the date-1 incentive constraint (2) is violated under every experimentation policy.

Henceforth, we assume that condition (10) holds to focus on non-empty optimal relational contracts.

Step 1: Any non-empty optimal contract implements FB, GE or GW. Since condition (10) holds, the players can always implement the first-best in state $n = 2$ by permanently exploiting both projects. At the other extreme, when $n = 0$, the players choose between exploring one or both domains.

Two explorations when $n = 0$: The players can explore both domains in period $n = 0$ if and only if they can implement FB. Necessity is immediate: FB prescribes two simultaneous explorations at $n = 0$. For sufficiency, assume the players are indeed able to carry out two explorations at $n = 0$. If a project with value $\bar{v}$ is discovered in one domain ($n = 1$), the action profile (exploit, explore) must also be implementable. The expected continuation surplus under (explore, explore) at $n = 0$ is strictly lower than under (exploit, explore) at $n = 1$; hence Inequality (2) is satisfied in the first case and slack in the second. Consequently the entire FB path is feasible, completing the equivalence.

One exploration when n=0: At $n = 1$, if the players are active in both domains, Observations 1 and 2 above force the GE policy. If only one domain remains active, the players either adopt the GW policy or pursue an alternative policy consisting of permanently exploiting the single high-value project found to date. The latter policy cannot be optimal: simply replicating the same exploration/exploitation rule in the other domain would strictly raise surplus (c.f., Bernheim and Whinston, 1990). Nor can scope drop to zero at $n = 1$: reverting to the $n = 0$ policy always yields strictly higher surplus.

Step 2: Surplus Ordering. The claim that FB yields strictly greater surplus than GE, which in turn yields strictly greater surplus than GW, follows from the fact that exploiting as many $\bar{v}$ projects as early as possible strictly increases surplus.

Step 3: Threshold Ordering. Recall that δ^0 denotes the threshold above which FB is implementable under an optimal relational contract. We first show that $\delta^{GE} < \delta^0$ for all parameter values.

In what follows, we focus, for each policy, on the on-path histories where the incentive constraints implied by Inequality (2) are most binding, since a policy that satisfies these constraints must be implementable by an optimal relational contract everywhere else on path. For the FB policy, the most stringent constraint arises when $n = 0$:

$$2c \leq 2\left(B^0(0) - \mathbb{E}(v_p - 2c)\right) \iff c \leq \left(B^0(0) - \mathbb{E}(v_p - 2c)\right), \tag{11}$$

where $B^0(0) - \mathbb{E}(v_p - 2c)$ is the definition of the continuation value under FB. Notably, δ

increases $B^0(\cdot)$ continuously, and thus, as shown in Proposition 3, we have $0 < \delta^0 < 1$.

Further, suppose the players implement GE, beginning with, say, exploration in domain 1. There exists an optimal relational contract that implements this policy if and only if:

$$c \leq \Big(B^0(0) - \mathbb{E}(v_p - 2c)\Big) + \frac{\delta \mathrm{Pr}\left(v_p = \bar{v}\right)}{1 - \delta \cdot \mathrm{Pr}\left(v_p \neq \bar{v}\right)} B^0\left(0\right), \tag{12}$$

$$2c \leq \frac{\delta}{1-\delta}\left(\bar{v} - 2c\right) + B^0\left(0\right) - \mathbb{E}\left(v_p - 2c\right). \tag{13}$$

Inequality (12) governs the case $n = 0$, when both domains are unexplored: the players must wait one discounted period to discover a $\bar{v}$ project in the first domain before turning to the second. Inequality (13) applies at $n = 1$, when they exploit one $\bar{v}$ project while exploring the other domain. Since both constraints depend continuously on δ, it suffices to check slackness at δ^0 to conclude $\delta^{IE} < \delta^0$. Slackness of (12) follows directly from $B^0(0) > 0$. Slackness of (13) follows by combining (11) with condition (10), which itself holds strictly. This completes this step of the proof.

Finally, to show that GW can be optimal, it suffices to find values of $(c, \underline{v}, \bar{v}, \mathrm{Pr}(v_p = \bar{v}))$ for which all of the necessary inequalities for $\delta^{GW} < \delta^{GE}$ hold strictly. This is sufficient as each of these inequalities is continuous continuous with respect to $(c, \underline{v}, \bar{v}, \mathrm{Pr}(v_p = \bar{v}))$. The result is established by computation (alongside the other claim that $\delta^{GW} > \delta^{GE}$ also occurs), and the code is available upon request. □

In the binary-value setting, an initially limited experimentation policy is always optimal for intermediate discount factors. Because project values are binary, there is no efficiency gain from pooling incentives across domains in the early phases of an initially maximal policy–the distribution fails the "richness" requirement of Definition 1. Without cross-domain incentive-pooling, the sole mechanism that can undermine the optimality of initially limited policies in the general case (see Proposition 5) vanishes.

Moreover, when $\delta \in [\delta^{\mathrm{GW}}, \delta^{\mathrm{GE}})$, players display one of the two behaviors described in Proposition 4: they may recall projects previously abandoned. Providing intuitive conditions for GW's optimality is delicate because it hinges on two opposing effects. On the one hand, GW dampens the players' temptation to renege at $n = 1$; on the other, it postpones exploitation of a discovered $\bar{v}$ project. Both forces feed back into the incentive constraint (2) at $n = 0$, tightening or loosening it depending on the precise parameter values.

5.2 Comparative Statics of Scope

The maximum potential scope of experimentation, m, can vary significantly depending on the application. When firms pool resources, some pairings may yield numerous cooperation opportunities, while others result in fewer viable collaborative areas, depending on the complementarity of their assets. In this subsection, we analyze how variations in m affect the profitability and sustainability of experimentation.

Before proceeding, we revisit Bernheim and Whinston (1990)'s analysis of scope, in stationary environments without learning dynamics. First, for a scaling factor $k \geq 1$, when scaling the scope of interaction by k, players can maintain the same per-domain average payoffs by replicating the original equilibrium k times independently. Second, when domains are identical, pooling incentives across domains cannot improve the players' per-domain average payoffs. However, if domains are asymmetric, players may gain from doing so and, hence, greater scope may be beneficial.

Let $\tilde{\pi}(m) := \pi(m)/m$ denote the average joint surplus per domain of the collaboration. Recall that $\delta^*(m)$ represents the minimum discount factor for which the optimal relational contract is non-empty. For a scaling factor $k \geq 1$, the following weak inequalities follow from Bernheim and Whinston (1990): $\tilde{\pi}(mk) \geq \tilde{\pi}(m)$ and $\delta^*(mk) \leq \delta^*(m)$.[19] In our setting, we can provide necessary and sufficient conditions for these inequalities to hold strictly, due to the dynamics stemming from the players' exploration of projects. Specifically, $0 < \delta^*(m \cdot k) < \delta^*(m)$ for $k > 1$ if $(1-\alpha)\mathbb{E}(v_p) < c$ and otherwise $\delta^*(m \cdot k) = 0$

[19] $\tilde{\pi}(m)$ is not necessarily monotone in m. For example, it may depend on the parity of m—pooling incentives across two domains could enable a relatively efficient experimentation policy, yet leave insufficient slack to improve efficiency in a third domain (as seen in the distribution used to prove Statement 2 of Proposition 6).

regardless of k. When $(1-\alpha)\mathbb{E}(v_p) < c$, the optimal relational contract will be empty for low discount factors. In these instances, scaling up m will strictly decrease δ^*. To see why, note that if the players were to implement k independent and concurrent collaborations, each with an identical experimentation policy, the threshold $\delta^*(m \cdot k)$ would be independent of k. However, this approach would be inefficient as it only leverages relational interdependencies within segmented multi-domain experimentation policies. Therefore, the players could sustain a non-empty relational contract for lower discount factors by leveraging interdependencies across all $m \cdot k$ domains. By an identical reasoning, $\tilde{\pi}(m \cdot k) > \tilde{\pi}(m)$ whenever the second-best experimentation policy is non-empty.

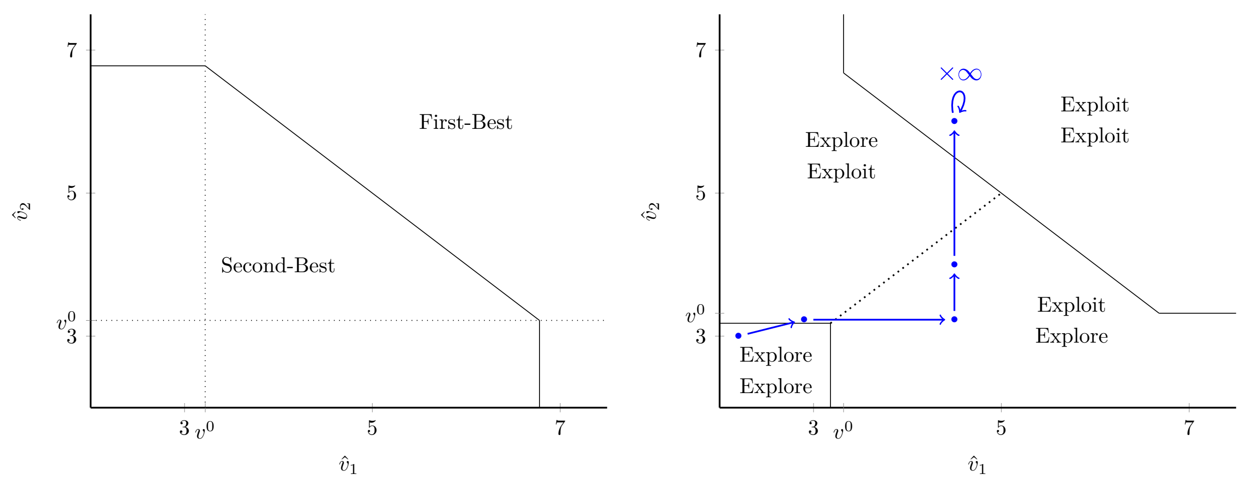
5.3 Multi-Project Collaborations: A Graphical Illustration

We analyze an example with specific parameter values. We set $c = 1$, $\delta = 1/3$, $\alpha = 1/2$, and $m = 2$. Finally, the project values v_p are drawn from a shifted exponential distribution with a rate parameter $\lambda = 1/2$, i.e., $v_p \sim 1 + \text{Exp}(1/2)$. Under this distribution, $\mathbb{E}(v_p) = 3$. The players' scope of experimentation is always maximal since $\alpha\mathbb{E}(v_p) - c = (1-\alpha)\mathbb{E}(v_p) - c > 0$, making exploration preferable to being idle. Further, the continuation value $\mathcal{C}(\hat{v}_1, \hat{v}_2)$ is weakly greater than 1 for all $\hat{v}_1$ and $\hat{v}_2$, as players can always explore two new projects per period, yielding a payoff of $\mathbb{E}(v_p) - 2c = 1$ per project and a continuation value $\mathcal{C}(\hat{v}_1, \hat{v}_2)$ also equal to 1. As a result, if Inequality (5) does not hold, players either: (i) exploit one project while exploring another, or (ii) explore two projects simultaneously.

Figure 1a. The figure depicts the first-best policy stated in Proposition 1. The vertical and horizontal black dotted lines represent the time-invariant threshold v^0 for domains 1 and 2, respectively. In both domains, projects with values above this threshold are permanently exploited, while those below are never exploited.

Further, the solid black line in the figure divides the project value space into two distinct regions. This line represents the set of $(\hat{v}_1, \hat{v}_2)$ values satisfying $h(\hat{v}_1, \hat{v}_2) = \tilde{v}$, a condition stated in Proposition 3. To the northeast of this line, in the region labeled "First-Best," the players can implement the first-best experimentation policy. To the southwest of the line, in the region labeled "Second-Best," the players can exploit at most one project at a time.

The horizontal segment represents where $\hat{v}_1 < v^0$, so project 1 is never exploited under the first-best policy, and implementation depends solely on $\hat{v}_2$. Symmetrically, the vertical segment shows where $\hat{v}_2 < v^0$, with implementation depending only on $\hat{v}_1$. The downward-sloping segment captures instances where both $\hat{v}_1$ and $\hat{v}_2$ exceed v^0. Here, increasing one project's value allows decreasing the other's while maintaining sufficient continuation value for first-best policy implementation.



(a) Feasible Region for First-Best Experimentation

(b) Project Exploitation and Sample Path

Figure 1: Optimal Multi-Project Experimentation

In the figure, we assume $c = 1$, $m = 2$, $\delta = 1 \,/\, 3$, and $v_p \sim 1 + \text{Exp}(1 \,/\, 2)$. $\hat{v}_1$ and $\hat{v}_2$ denote the values of the best projects discovered in domains 1 and 2, respectively. The left figure plots (i) the threshold v^0 for switching from exploration to exploitation in the first-best and (ii) the set of $\hat{v}_1$ and $\hat{v}_2$ values satisfying $h(\hat{v}_1, \hat{v}_2) = \tilde{v}$ in solid black. The right figure divides the project value space into four regions, determined by the exploitation or non-exploitation of each project. The top mention indicates the decision for the project with value $\hat{v}_1$, while the bottom mention shows the decision for the project with value $\hat{v}_2$. In Blue, we plot one realization of a sample path of project values.

Figure 1b. The project value space is divided into four regions, determined by the exploitation or non-exploitation (in favor of exploration) of each project. The top mention indicates the decision for the project with value $\hat{v}_1$, while the bottom mention shows the decision for the project with value $\hat{v}_2$. It follows from Figure 1a that both projects are chosen for exploitation when in the "First-Best" region and $\hat{v}_1, \hat{v}_2 \geq v^0$. Outside of this region, the players can choose one project for exploitation at most. One can prove that there exists a threshold, v', on the value of the best of the two projects such that, below this threshold, the players choose to explore two new projects rather than exploiting the best of the two projects. The threshold v' corresponds to the boundary of the "Explore Explore" region in the figure. We observe that the threshold v' is lower than v^0, indicating that players may opt to exploit a project even when they are certain to not permanently exploit it in the future.[20]

Figure 1b also presents a sample path illustrating the evolution of $(\hat{v}_1, \hat{v}_2)$ over time, depicted in blue. In the early phase where the players are exploring two projects, both $\hat{v}_1$ and $\hat{v}_2$ weakly increase over time. In the phase where the players exploit a project in domain j, $\hat{v}_j$ remains constant, while $\hat{v}_{-j}$ weakly increases over time. Finally, in the phase where the players exploit both projects, $\hat{v}_1, \hat{v}_2$ stay constant because exploitation is permanent. Arrows are used to signify changes in project values when a more valuable project is identified, while self-loops indicate situations where more valuable projects are either not discovered or not pursued. The path shown in the figure includes temporary exploitation in domain 2 (of a project guaranteed to be not permanently exploited), as discussed in Proposition 4.

5.4 Beyond Independent and Identical Domains

Our main analysis assumed identical and independent collaboration domains. In practice, firms often collaborate across domains with diverse characteristics and technological interdependencies. This reality raises the question: Which domains, if any, should be prioritized when initiating collaboration? The Online Appendix explores three scenarios that address these questions and formulate predictions. We briefly summarize these extensions here.

When to explore risky domains?

Our main analysis, by assuming an infinite number of independent and identically distributed

[20] The threshold v' is computed using numerical integrals and approximate solutions to the Bellman equation. The result that v' can be lower than v^0 can be proven analytically.

projects, effectively eliminated risk considerations. However, collaborating parties often face uncertainty about their collaboration's potential value, with varying degrees of uncertainty across cooperation domains. For instance, a buyer-supplier collaboration might involve both incremental improvements to an existing product and the development of a radically new—and thus potentially unprofitable—project. To capture these features, we modify a two-domain version of our framework by supposing that that the first domain is exactly as in the main model, while the other contains a single project with either low or high value. We show that even when immediate cooperation across both domains is feasible, players may choose to postpone exploring the risky domain 2 project. This delay continues until a sufficiently valuable project is discovered in domain 1. Such a gradual approach safeguards the collaboration against complete dissolution should the radical innovation fail.

Can "win-win" projects serve as stepping-stones?

In the main analysis, we assumed that each project's benefits accrue to only one player. However, the model can be extended to reflect more nuanced real-world scenarios. Collaborating parties often engage in both "win-win" projects yielding mutual benefits and projects that disproportionately advantage certain participants. In modeling these scenarios, this extension assumes two domains with distinct benefit structures. In one domain, projects yield equal benefits to both players.[21] The other domain follows the main analysis, where project benefits accrue exclusively to one player. We show that optimal experimentation is initially limited for low values of the discount factor and that the domain with symmetric projects is explored first.

How do technological interdependencies influence gradualism?

In the third extension, we introduce positive correlation between project values across domains, such that discovering a valuable project in one domain immediately reveals a project of equal value in the other. This assumption reflects how success in one area can enhance

[21] The Online Appendix includes another extension in which domains differ in the probability α, with qualitatively similar results.

opportunities in another (e.g., mRNA technology's wide applicability across medical conditions). Absent incentive issues, players would optimally explore both domains concurrently to expedite valuable project discovery. With asymmetric benefits, an initially limited approach is strictly optimal for intermediate discount factors. These findings suggest initially limited approaches are more likely to be optimal in R&D environments with stronger cross-domain knowledge spillovers.

6 Applied Insights

This section connects our theoretical analysis to two key literatures: buyer-supplier relationships and persistent productivity differences across firms. In Online Appendix Section 3 we return to some of the models surveyed in Section 1.1, comparing how our framework explains the qualitative patterns documented here relative to these alternative approaches.

6.1 Buyer-Supplier Collaborations

The economics literature on buyer-supplier relationships has largely focused on vertical integration with relationship-specific investments (Williamson, 1975; Grossman and Hart, 1986; Hart and Moore, 1990), optimal contracts under externalities or agency frictions (see Tirole, 1988, Chapter 4), and more recently, relational contracts for supplier allocation (Board, 2011; Andrews and Barron, 2016). These studies typically assume predetermined gains from trade. In contrast, our work examines settings where such gains must be discovered through collaborative experimentation, often spanning multiple products or markets.

We formalize that discovery process in a deliberately stylized framework, stripping away technological and institutional features such as cross domain spillovers or cash flow constraints, so that all dynamics stem from strategic interaction alone. Although these omitted features matter in practice and therefore limit the model's ability to reproduce every real-world pattern, the strategic forces we highlight should still operate alongside them. In our set-up the parameter m denotes the number of product categories or market geographies. Both firms make non-contractible investments of c, which are observable but not verifiable by third parties. Innovation requires complementary, indispensable inputs from both sides. Even after the exploration phase, further non-contractible investments (also c) remain necessary—for example, worker training or marketing. Benefits are asymmetric: final proceeds accrue to the buyer (high α), who compensates the supplier via upfront transfers w or bonuses b.

Our theoretical analysis both draws on and contributes to a large case-study literature on buyer-supplier dynamics. This work highlights experimentation and trust as key to successful collaboration, especially when benefits are distributed asymmetrically. A McKinsey report highlights this asymmetry of benefits: "Some collaborations promise equal benefits for both parties. [...] In other cases, however, the collaboration might create as much value overall but the benefit could fall more to one partner than to the other" (Benavides et al., 2012). Such asymmetries underscore the central role of trust, given the limits of formal contracts. Doney and Cannon (1997) distinguish between two types of trust: "benevolence" trust (belief in a partner's genuine desire to collaborate) and "credibility" trust (expectation that a partner will fulfill promises due to self-interest). Our analysis focuses on credibility trust, operating under the assumption that both parties desire collaboration. Consequently, we emphasize work that similarly concentrates on credibility trust. The concept of benevolence trust, while important, corresponds more directly to the analyses by Watson (1999, 2002), which we discuss in Section 1.1.

Dwyer et al. (1987) highlight the dynamic nature of buyer-supplier relationships, emphasizing the central role of relational contracts. They describe an initial "search and trial phase" that evolves into an "expansion phase," characterized by increased risk-taking and deeper mutual dependence. As they note, "The rudiments of trust and joint satisfactions established in the exploration stage now lead to increased risk taking within the dyad. Consequently, the range and depth of mutual dependence increase." A senior executive from a

Toyota supplier similarly described their relationship with Toyota: "We started by making one component, and as we improved, [Toyota] rewarded us with orders for more components" (Liker and Choi, 2004). The common pattern of these relationships starting small before expanding is consistent with our findings, particularly Proposition 5. It also supports our extension in Section 5.4, which examines the strategic delay of high-risk ventures in these relationships.[22]

Building on Dwyer et al. (1987), Vanpoucke et al. (2014) corroborate both the prevalence of gradualism and the occurrence of extended experimentation periods in buyer-supplier relationships. These phenomena are driven by the parties' need to establish credibility in the context of relational contracts. As one CEO in their study noted, "We use contracts, but not everything, certainly in the long run, can be put in contracts." Their case study of soybean product development, where partners took a decade to initiate integration and build sufficient credibility, illustrates this phenomenon. This evidence is consistent with our analysis, particularly Corollary 1, which predicts that collaborating firms must engage in prolonged experimentation in order to identify joint projects of sufficient value to sustain the subsequent exploitation phase. Furthermore, Vanpoucke et al. (2014) emphasize the strong path dependence of relationship dynamics, observing that "events, rather than time," define relationship development stages. Their case studies consistently reveal that successes in initial cooperation domains typically drive further joint collaborations. This supports our model, where scope expands through discrete "events" that shift the players' continuation value, rather than through the mere passage of time.

Lastly, Proposition 6 showed that the long-term scope of a collaboration is determined during the initial phases, with early outcomes influencing the trajectory and ultimate ex-

[22]Macchiavello and Morjaria (2015) show that in the context of the Kenyan Rose export market, the duration of buyer-supplier relationships is predictive of the scale of the relationship. Moreover, the trade literature has documented the strong persistence of buyer-supplier relationships, ascribed to substantial switching or search costs (see e.g., Bernard and Moxnes, 2018; Monarch, 2022). This literature has also noted the tendency for transaction volumes to increase over long periods of time (see e.g., Monarch and Schmidt-Eisenlohr, 2023). Our framework suggests that buyers may exhibit reluctance to switch suppliers when such a transition entails the need to rebuild a relationship from scratch.

tent of the partnership. Dwyer et al. (1987) characterize the early exploration phase in buyer-supplier relationships as "very fragile," highlighting the critical nature of these initial interactions. Benavides et al. (2012) provide a concrete example of this fragility, describing a case where an early collaboration attempt between a retailer and manufacturer yielded somewhat disappointing results. While their relationship did not terminate entirely, Benavides et al. (2012) suggest that this initial setback was the primary reason their partnership did not expand further. Proposition 6 also predicts that collaboration scopes can contract over time without collapsing entirely. For example, in 2015 Ericsson and Cisco unveiled the "Networks-of-the-Future" alliance, pledging a joint portfolio that spanned routing, RAN, Wi-Fi, cloud, and global services in more than 180 countries (Cisco Systems Inc., 2015). By mid-2017 Ericsson's CEO acknowledged that the collaboration had produced only a handful of joint wins (Marek, 2017). Analysts noted that front-line sales teams tended to "sell their own gear first," which made sustained cooperation difficult (Gonsalves, 2017). Confronted with these incentive problems, the partners decreased the alliance to a few clearly defined product combinations and legacy service contracts (Matsumoto, 2017).

6.2 Persistent Performance Differences

While our focus has been on firm-to-firm interactions, the model also offers a useful lens for employer-employee dynamics. One party can be viewed as the employer, the other as the employee, with benefits, for example, accruing primarily to the employer. The different domains then correspond to distinct dimensions of the production improvement process.

With this interpretation, our work also contributes to the literature on persistent performance differences among seemingly similar firms (see Syverson, 2011; Gibbons and Henderson, 2013, and references therein). Numerous empirical studies document durable performance gaps across industries—with these gaps proving robust against plausible explanations such as market competition or local geographical and demand conditions, and are instead strongly linked to managerial practices (c.f. Bloom and Van Reenen, 2007). According to

Gibbons and Henderson (2013) and the evidence they review, differences in managerial practices—rooted in the nature of relational contracts—are central to explaining productivity disparities. We adapt their categorization of explanations: (i) managers may be unaware of their poor performance or view external best practices as inapplicable; (ii) they are aware but choose not to adopt better practices; or (iii) they are "striving mightily" to implement improvements but face obstacles in doing so.

The first explanation underscores information barriers, prompting questions about why such information does not diffuse more readily (c.f. Bloom et al., 2013; Atkin et al., 2017). The second explanation is consistent with the framework developed by Chassang (2010) and discussed in Section 1.1, in which players are informed about the existence of more efficient practices but choose not to pursue them to preserve their relationship.[23] Our analysis in Section 4.3 offers a complementary rationale for explanation (ii), showing that long-run collaboration may remain inefficiently limited. Transitioning from exploration to exploitation in one domain can reduce the ability to cooperate in others, leading to constrained scope.

Unlike other models we know, our model also offers insight into explanation (iii) presented by Gibbons and Henderson (2013). Consider two organizations with identical characteristics implementing ex-ante identical experimentation policies, operating under a discount factor where the scope of experimentation is initially limited. Their paths diverge if one organization discovers a highly valuable practice early on, thus expanding its scope, while the other does not. The second organization, still attempting to achieve any success, appears to be "striving mightily" to match the first organization's performance. However, identifying superior practices is time-intensive. The second organization cannot increase its scope until it finds a sufficiently valuable practice, potentially leading to a persistent performance gap.

[23]Chassang (2010) establishes the possibility of arbitrarily small persistent productivity differences in the binary-value case. However, unlike in our setting, it cannot generate substantial long-run differences of the kind documented in the empirical literature, as it does not fully characterize the optimal punishment phases. As a result, it cannot compare the expected payoffs of players with meaningfully different histories of play.

7 Concluding Remarks

This paper presents a framework for analyzing the dynamics of multi-domain collaborative experimentation when benefits are unevenly distributed and policies must be self-enforcing. Our model yields four key insights. First, when the initial relationship value is low, parties do not treat each domain independently and engage in extended exploration phases. Second, cross-domain relational interdependence leads to seemingly counterintuitive exploration/exploitation decisions, including prolonged exploitation of ultimately discontinued projects or revival of abandoned ones. Third, experimentation may progress gradually: parties may begin by exploring some domains and expand to others based on early success, with some domains possibly left unexplored. Fourth, even if all domains are initially explored, the players may end up exploiting projects in only a limited number of domains.

While our main focus is on buyer-supplier dynamics and firm productivity, the framework also applies to political economy settings involving multi-domain collaboration. In federal systems, central governments use fiscal transfers to encourage subnational policy experimentation (c.f. Callander and Harstad, 2015; Wang and Yang, forthcoming). Our analysis underscores strong path dependence in policy implementation and points to unexpected cross-domain spillovers. Similarly, political and economic unions like the EU promote collaboration via shared resources and structural funds, while allowing members to exit. Consistent with our model, EU integration proceeded gradually—starting with mutually beneficial projects and later expanding to costlier, unevenly distributed policies (see, e.g., Spolaore, 2015, and references therein, as well as Section 5).

Future work could extend the framework in several directions. Relaxing the assumption of i.i.d. project benefits within domains would allow exploration of directed innovation strategies and distinctions between radical and incremental innovation (c.f. Callander, 2011; Garfagnini and Strulovici, 2016; Callander and Matouschek, 2019). We also assumed both players are required for exploration and exploitation, with outside options fixed. Future research could examine how evolving outside options—shaped by experimentation history—affect collaboration. Finally, introducing asymmetric roles offers a natural extension. For example, exploration might involve only one player (e.g., R&D), while exploitation depends on another (e.g., Sales), allowing analysis of settings where these tasks are disentangled (see Krieger et al., 2019; Lizzeri et al., 2024, for qualitative and theoretical treatments).

Appendix

Proof of Proposition 1. Following a reasoning almost identical to that in Bergemann and Välimäki (2001), player 0 treats each domain independently and identically and never recalls a project because $|\mathcal{P}_j| = \infty \ \forall j$. Therefore, the optimal policy conditions only on the project with the highest value amongst all previously explored projects, whose value we denote $\hat{v}$. The Bellman equation for player 0 can be defined for any value $\hat{v}$—even if $\hat{v}$ is not in the support of v_p—and is given by:

$$B^0(\hat{v}) = \max_{\text{explore, exploit } \hat{v}} \left\{ \mathbb{E}\left(v'\right) - 2c + \delta \mathbb{E}\left(B^0\left(\max\left(\hat{v}, v'\right)\right)\right), \hat{v} - 2c + \delta B^0\left(\hat{v}\right) \right\}. \tag{14}$$

The first term in the maximum operator corresponds to the player's expected surplus when exploring one more project and the second term is their surplus when exploiting the project with value $\hat{v}$. Next, there exists a threshold v^0, wherein the players explore if $\hat{v} < v^0$ and exploit if $\hat{v} \geq v^0$. Further, Blackwell's Sufficient Conditions imply that there exists a unique solution to the Bellman equation, and hence the threshold rule dictated by v^0 is a solution. This threshold is determined by:

$$\frac{1}{1-\delta}(x - 2c) = \mathbb{E}(v_p - 2c) + \frac{\delta}{1-\delta}\mathbb{E}(\max\{v, x\} - 2c). \tag{15}$$

If $x \in \text{Supp}(v_p)$, then $v^0 = x$. If $x \notin \text{Supp}(v_p)$, then $v^0 := \inf\{y : y \geq x \text{ and } y \in \text{Supp}(v_p)\}$ which always exists in the closure of $\text{Supp}(v_p)$. Finally, standard comparative statics arguments imply that v^0 is increasing in δ. $\square$

Proof of Proposition 2. Recall that after a deviation in period t, players set $P_i^t = \emptyset$ and $b_{i,-i}^t = 0$ if not already chosen. In subsequent periods, they revert to the static equilibrium with zero transfers and no selected projects.

The proof proceeds in four steps: (i) we show that it is without loss of optimality to restrict attention to relational contracts that are surplus-maximizing following every on-path history h^t; (ii) we provide a necessary and sufficient condition for the existence of a relational contract that implements a given experimentation policy $\hat{\mathbf{P}}(\cdot)$; (iii) we show that this condition is independent of the division of surplus between the players; and (iv) we show that, for any two histories that generate the same beliefs, selecting the same continuation equilibrium is without loss of optimality.

Step 1 We show that it is without loss of optimality to restrict attention to relational contracts that are surplus-maximizing following every on-path history h^t. To see this, suppose that there exists an on-path history h^t such that the continuation equilibrium starting in period t, denoted by e^1, has lower total surplus than an alternative continuation equilibrium e^2. Thus, if we define $\mathcal{C}_i^k$ to be the continuation value to player i in equilibrium e^k, then $\sum_i \mathcal{C}_i^1 < \sum_i \mathcal{C}_i^2$. For the rest of Step 1, we omit the superscript $t-1$ in our notation, as we are solely concentrating on period $t-1$ objects.

Let us modify the players' relational contract such that play in and after period t is dictated by e^2 and the period $t-1$ $b_{i,j}(\cdot)$ transfers associated with history h^t (and, thus, corresponding to a specific realizations of $\mathbf{x}^{t-1}, \mathbf{v}^{t-1}$) are adjusted so that: (i) player 2's expected payoff following the realizations of $\mathbf{x}^{t-1}, \mathbf{v}^{t-1}$ is the same as under the original equilibrium and (ii) player 1's expected payoff following the realizations of $\mathbf{x}^{t-1}, \mathbf{v}^{t-1}$ increases by $\sum_i \mathcal{C}_i^2 - \sum_i \mathcal{C}_i^1$. Specifically, take the vector of transfers $\mathbf{b}_1 = (b_{1,2}^1, b_{2,1}^1)$ associated with the original equilibrium and create a new vector of transfers $\mathbf{b}_2 = (b_{1,2}^2, b_{2,1}^2)$ such that:

$$\mathcal{C}_1^2 + b_{2,1}^2 - b_{1,2}^2 > \mathcal{C}_1^1 + b_{2,1}^1 - b_{1,2}^1, \tag{16}$$

$$\mathcal{C}_2^2 + b_{1,2}^2 - b_{2,1}^2 = \mathcal{C}_2^1 + b_{1,2}^1 - b_{2,1}^1. \tag{17}$$

Because $\sum_i \mathcal{C}_i^2 - \sum_i \mathcal{C}_i^1 > 0$, finding payments that satisfy $b_{1,2}^2 \leq \mathcal{C}_1^2$ and $b_{2,1}^2 \leq \mathcal{C}_2^2$ is always feasible.

Note that these changes have no impact on player 1's choices of actions made in any period $t' \leq t-1$ because all actions are observable, and hence choosing a different action from the proposed equilibrium would be labeled a defection. If defections were deterred in the original equilibrium, which had a strictly smaller continuation value for player 1, then they are also deterred in the new equilibrium. The same logic applies to player 2 since they obtain the same expected payoff in period $t-1$ (compared to the original equilibrium), and thus also have the same continuation values in all periods $t' < t-1$. Finally, note that surplus from a date 0 perspective is strictly higher under the new equilibrium.

Step 2 We show that there exists a relational contract that implements an experimentation policy $\hat{\mathbf{P}}(\cdot)$ if and only if the following inequality holds for all t and for all histories $h^t \in \mathcal{H}^t$:

$$\sum_{p \in \hat{\mathbf{P}}^t} \sum_{i=1,2} \max\Big(0, c - \mathbb{E}(v_p \mathbb{1}_{x_p=i} | h^t)\Big) \leq \mathcal{C}(h^t), \tag{18}$$

where $\mathcal{C}(h^t)$ is the continuation value.

To show that (18) is a necessary and sufficient condition, consider a set of transfers $b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \geq 0$ to be paid on path given a vector of realized values $\mathbf{x}^t, \mathbf{v}^t$.

Given an equilibrium experimentation policy $\mathbf{P}^t$, note that it is without loss of generality to assume that $P_1^t = P_2^t = \mathbf{P}^t$. Thus, for each player and for each $p \in \mathbf{P}^t$, the player must weakly prefer to include p in P_i^t, rather than excluding it. Let $\sigma_i(\mathbf{x}^t, \mathbf{v}^t)$ denote player i's share of $\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)$ as a function of $\mathbf{x}^t, \mathbf{v}^t$. Hence, the condition for selecting $\mathbf{P}^t$ is:

$$\begin{aligned}\sum_{p \in \mathbf{P}^t} \max\Big(c - \mathbb{E}(v_p \mathbb{1}_{x_p=i} | h^t), 0\Big) \leq \mathbb{E}\Big(& b_{-i,i}(\mathbf{x}^t, \mathbf{v}^t) - b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \\ &+ \sigma_i(\mathbf{x}^t, \mathbf{v}^t) \mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)\Big), \quad \forall i,\end{aligned} \tag{19}$$

$$b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \leq \sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t\mathbf{x}^t, \sqcup\mathbf{v}^t), \quad \forall\mathbf{v}^t, \forall i. \tag{20}$$

Expectations are taken over the project valuations realizations $\mathbf{x}^t, \mathbf{v}^t$ and $h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t$ denotes the players' updated beliefs after observing $\mathbf{x}^t, \mathbf{v}^t$.[24] The first expression states that the promised transfers and the expected share of the total continuation value must be enough to prevent a player from shirking on any subset of the projects. The second expression states that each player is willing to pay the other player the necessary transfer.

To show necessity: Note that since Equation (19) must hold for a fixed i, the inequality also holds summing over all i. Further, all transfers cancel out when summing over i. Finally, by definition, $\mathbb{E}(\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)) = \mathcal{C}(h^t)$. Hence, we are left with Equation (18).

To show sufficiency: We will show this result in two substeps.

SubStep 1: We show it is necessary and sufficient to replace Equation (20) by its expectation. This new expression is as follows:

$$\mathbb{E}(b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)) \leq \mathbb{E}\Big(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)\Big) \quad \forall i. \tag{21}$$

We first show that if there is a solution to Equations (21) and (19), then there exists a solution to Equations (20) and (19).

Take a set of transfers $b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)$ that satisfy Equations (21) and (19). Define:

$$\begin{aligned} b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) =& \sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t) \\ &- \Bigg(\mathbb{E}\Big(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t) - b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)\Big)\Bigg). \end{aligned} \tag{22}$$

Since Equation (21) holds, the term in the expectation of Equation (22) is positive and thus Equation (20) holds for all realizations of $\mathbf{x}^t, \mathbf{v}^t$ under the set of transfers $b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)$.

[24]The history also includes the project selections, and both the upfront and end-of-period transfers. However, for notational convenience we only include the realized valuations as every other object can be inferred on path from the realized valuations.

Finally, $\mathbb{E}(b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)) = \mathbb{E}(b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t))$ so Equation (21) continues to hold.

SubStep 2: Using substep 1, it suffices to show that Equation (18) implies a solution to Equations (19) and (21). To simplify all the notation with expectations, Equation (19) can be re-expressed as:

$$\beta_i - \gamma_i \leq (\tilde{b}_{-i,i} - \tilde{b}_{i,-i}), \tag{23}$$

where $\tilde{b}_{i,-i}$ is the expected transfer from i to $-i$, $\beta_i = \sum_{p \in \mathbf{P}^t} \max\Big(0, c - \mathbb{E}(v_p \mathbb{1}_{x_p=i} | h^t)\Big)$, and $\gamma_i = \mathbb{E}(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t))$. Equation (21) can thus be re-written as:

$$\tilde{b}_{i,-i} \leq \gamma_i. \tag{24}$$

Rearranging Equation (18) implies $\sum_i (\beta_i - \gamma_i) \leq 0$. One can now show that $\tilde{b}_{i,-i} = \max(0, \beta_{-i} - \gamma_{-i})$ satisfies Equation (24). Further, Equation (23) holds because:

$$\begin{aligned} \beta_i - \gamma_i \leq \max(0, \beta_i - \gamma_i) - \max(0, \beta_{-i} - \gamma_{-i}) && (25) \\ \iff \max(0, \beta_{-i} - \gamma_{-i}) - \min(0, \gamma_i - \beta_i) \leq 0 && (26) \\ \Longleftarrow \sum_i (\beta_i - \gamma_i) \leq 0, && (27) \end{aligned}$$

where the final step follows from noting that $\beta_1 - \gamma_1$ and $\beta_2 - \gamma_2$ cannot both be positive and analyzing the remaining three cases based on the signs of $\beta_i - \gamma_i$.

Finally, Equation (24) reduces to

$$\begin{aligned} \max(0, \beta_{-i} - \gamma_{-i}) \leq \gamma_i \Longleftarrow \beta_{-i} - \gamma_{-i} \leq \gamma_i && (28) \\ \Longleftarrow \sum_i (\beta_i - \gamma_i) \leq 0, && (29) \end{aligned}$$

where the final implication is due to β_i being weakly positive.

Step 3: We show that any relational contract that implements a given experimentation policy can be replaced by an alternative relational contract that implements the same experimentation policy and yields no surplus to player 2. First, note that the way the players share their continuation value does not affect Equation (2) from the main text. Hence, for any period t where player 2's expected payoff is positive, $w_{2,1}$ can be increased until player 2's expected payoff is zero. Player 2 is willing to make this transfer because not doing so would be seen as a deviation, resulting in a payoff of 0 for player 2.

Step 4: We now show that, for any two histories h_1^t and $h_2^{t'}$ that generate the same beliefs μ, selecting the same continuation equilibrium is without loss of optimality. Take a relational contract r that is surplus-maximizing at all on-path histories and has two histories h_1^t and $h_2^{t'}$ prescribing different (surplus-maximizing) continuation equilibria under the same beliefs μ. Recall from Step 3 that one can consider relational contracts in which player 2 obtains an expected payoff equal to 0 in every period. In this case, since the two continuation equilibria are both optimal and both give all the surplus to player 1, switching from one continuation equilibrium to the other does not change the players' incentives as both prescribe the exact same payoffs to the players. Hence, when focusing on relational contracts that specify the same continuation equilibrium following histories that induce the same beliefs, one can replace $\mathcal{C}(h^t)$ with $\mathcal{C}(\mu^t)$. □

Proof of Proposition 3. When the players have identified projects with values $\hat{v}_1, \ldots, \hat{v}_m$ at history h, the condition for the players being able to replicate the first-best experimentation policy in all subsequent periods is that, for all histories h' occurring after h and with associated project values $\hat{v}'_1, \ldots, \hat{v}'_m$, the players exploit $\hat{v}'_j$ if and only if $\hat{v}'_j \geq v^0$. This condition is as follows:

$$c \sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j \geq v^0} + \max\{0, c - (1-\alpha)\mathbb{E}\left(v_p\right)\} \sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j < v^0} \leq \sum_{j=1}^{m} \mathcal{C}^0(\hat{v}'_j), \tag{30}$$

$\forall (\hat{v}'_1, \ldots, \hat{v}'_m) \geq (\hat{v}_1, \ldots, \hat{v}_m)$, which corresponds to (2) when the players implement the first-best policy and where $\mathcal{C}^0(\hat{v}'_j)$ denotes the continuation value associated with domain

j under the first-best policy. Note that $\mathcal{C}^0(\hat{v}'_j)$ (i) is constant below v^0, (ii) is such that $\lim_{x\uparrow v^0} \mathcal{C}^0(x) > \lim_{x\downarrow v^0} \mathcal{C}^0(x)$ and (iii) is increasing above v^0. Given such properties, setting $\hat{v}'_j = \max\{\hat{v}_j, v^0\}$ both minimizes the right-hand side and maximizes the left-hand side of (30). Thus, an equivalent condition is:

$$m \cdot c \leq \delta\Big(\sum_{j=1}^{m} \frac{1}{1-\delta}(\max\{\hat{v}_j, v^0\} - 2c)\Big). \tag{31}$$

Finally, the existence of a threshold δ^0 was proven in the text. □

Proof of Corollary 2. The characterization of $v^*(\delta)$ follows from Corollary 1. The existence of δ^* follows an identical argument to that made in the proof of Proposition 5. □

Proof of Proposition 4. We first prove $\delta^* < \delta^0$. As argued in Section 5.2, $\delta^*(m) < \delta^*(1)$ for any $m > 1$, therefore, it suffices to show the result for $m = 1$. Suppose $\delta < \delta^0$ and consider the policy described in Corollary 2. By definition of $\tilde{v}$, it suffices to show that there exists an optimal relational contract that implements this policy in period 1–that is, one that satisfies Inequality (2), which is most binding when $\alpha = 1$ (as we henceforth assume). At $\delta = \delta^0$, this constraint is as follows: $c \leq \mathcal{C}^0(0)$. Given our assumption of richness, and thus that $v^0 \in \text{Supp}(v_p)$, we can re-write $\mathcal{C}^0(0)$, based on indifference at v^0:

$$\frac{v^0 - 2c}{1-\delta} = \mathbb{E}(v_p - 2c) + \mathcal{C}^0(0, \text{explore}) \implies \mathcal{C}^0(0, \text{explore}) > \frac{\delta}{1-\delta}(v^0 - 2c). \tag{32}$$

Further, δ^0 is defined by: $c = \frac{\delta}{1-\delta}(v^0 - 2c)$. Therefore, for δ slightly below δ^0, the policy in period 1 satisfies Inequality (4) if and only if the continuation value is continuous at δ^0, which holds since $\tilde{v}$ is continuous with respect to δ and v_p satisfies richness, ensuring that $\Pr\left(v_p \in (\tilde{v}(\delta^0), \tilde{v}(\delta))\right)$ continuously approaches 0 as $\delta \uparrow \delta^0$.

To show the second part of the proposition, denote $t(p) = \inf_t\{t : p \in \mathbf{P}^t\}$. By contradiction, $\forall p \in \mathcal{P}$, either (i) $p \in \mathbf{P}^t \ \forall\ t > t(p)$ or (ii) $p \notin \mathbf{P}^t \ \forall\ t > t(p)$. Further, as exploitation is permanent, and the value from exploitation is monotone in the value of the project then, for each domain j, there exists a threshold $v_j^*(\hat{\mathbf{v}}_{-j})$ such that the players exploit a project with value $\hat{v}_j$ if and only if $\hat{v}_j \geq v_j^*(\hat{\mathbf{v}}_{-j})$, where $\hat{\mathbf{v}}_{-j}$ denotes the values of the best projects found in the remaining domains.

Note that $v_j^*(\cdot)$ is weakly increasing in each of its arguments; otherwise, with positive probability, statement 2 of the proposition would be satisfied. Further, if $\hat{\mathbf{v}}_{-j}$ is point-wise strictly greater than $\tilde{v}$, then Proposition 3 implies that $v_j^*(\hat{\mathbf{v}}_{-j}) < \tilde{v}$. Therefore if v_p satisfies richness at δ, then the players permanently exploit projects with value less than $\tilde{v}$, deriving a contradiction. □

Proof of Proposition 5. We first prove the existence of δ^*. Suppose $\delta_1 < \delta_2$ and, by contradiction, that the optimal experimentation policy is non-empty for δ_1 but empty for δ_2. The optimal experimentation policy for δ_1 yields strictly positive surplus and yet cannot be implemented at δ_2. However, holding fixed the policy, the left-hand side of (2) is independent of δ and the right-hand side is increasing in δ, implying that the experimentation policy is feasible under δ_2, which is a contradiction. This reasoning implies that a threshold exists. Finally, $\delta^* < 1$ since $\mathcal{C}(\cdot) \to \infty$ as $\delta \to 1$.

We now prove the existence of $\bar{\delta}$. Scope is initially maximal if and only if $m \cdot \max\{0, c - (1-\alpha)\mathbb{E}(v_p)\} \leq \mathcal{C}(\cdot)$. However, by an identical argument as that in the preceding paragraph, the right-hand side of (2) is increasing in δ and the left-hand side of (2) is independent of δ. This implies the existence of a threshold on δ. Further, when $\delta \to 1$, $\mathcal{C}(\cdot) \to \infty$, implying maximal scope. As a result, $\bar{\delta} < 1$.

$\delta^* \leq \bar{\delta}$ because any initially maximal relational contract is non-empty. We now show that this inequality is strict when m is sufficiently large. Inequality (2) implies

$$m(c - (1-\alpha)\mathbb{E}(v_p)) \leq B\left(\mathbf{0}|\bar{\delta}(m)\right) - m\mathbb{E}\left(v_p - 2c\right). \tag{33}$$

Next, note that for any $x > \tilde{v}$, $\lim_{\delta \uparrow \bar{\delta}(m)} B\left(x, 0, \ldots, 0|\delta\right) \geq B\left(\mathbf{0}|\bar{\delta}(m)\right)$. We construct a suboptimal, initially limited policy in which, during period 1, players explore a single domain in search of a project with value strictly exceeding $\tilde{v}$. If they find such a project, they expand their scope; otherwise, they terminate their relationship. $\delta^*(m) = \bar{\delta}(m)$ implies that this policy cannot be implemented for $\delta < \bar{\delta}(m)$. Therefore:

$$c - (1-\alpha)\mathbb{E}(v_p) \geq \bar{\delta}(m)\Pr(v_p > \tilde{v})B(\mathbf{0}|\bar{\delta}(m)). \tag{34}$$

Combining (33) and (34) implies:

$$\begin{aligned} &\bar{\delta}(m)\Pr(v_p > \tilde{v})B(\mathbf{0}|\bar{\delta}(m)) \leq \frac{B(\mathbf{0}|\bar{\delta}(m))}{m} - \mathbb{E}(v_p - 2c) \\ \iff\ &B(\mathbf{0}|\bar{\delta}(m))\Big(\bar{\delta}(m)\Pr(v_p > \tilde{v}) - \frac{1}{m}\Big) \leq -\mathbb{E}(v_p - 2c). \end{aligned} \tag{35}$$

However, as shown in the text, $B\left(\mathbf{0}|\bar{\delta}(m)\right)$ diverges to infinity. Hence, $\bar{\delta}(m)\Pr(v_p > \tilde{v}) - \frac{1}{m}$ converges to zero. But since (i) $\bar{\delta}(m)$ remains bounded away from zero when $\mathbb{E}(v_p)(1-\alpha) < c$ and (ii) we assumed $\liminf \Pr(v_p > \tilde{v}(\bar{\delta}(m)) > 0$, we arrive at a contradiction. □

Proof of Proposition 6. We first show that our notion of terminal scope is well defined. If there exists a period t for which the players conduct no explorations, then $\mathbf{P}^{t'} = \mathbf{P}^t$ for all $t' \geq t$. By contradiction, if there exists an equilibrium path where $\liminf |\mathbf{P}^t| < \limsup |\mathbf{P}^t|$, the players must explore at least one project in each period t. However, for each exploration on any given domain, with positive probability the players discover a project with value exceeding $\tilde{v}$. Moreover, since exploration occurs only in domains with the lowest $\hat{v}$ values, all domains eventually yield a project exceeding $\tilde{v}$, implying that the first-best policy becomes implementable in all subsequent periods. As a result, $\liminf |\mathbf{P}^t| = \limsup |\mathbf{P}^t|$. This argument also shows that terminal scope equals m with positive probability.

Statement 1: We assume $v_p \in \{0, \underline{v}, \bar{v}\}$, $m = 2$, and $\alpha = 1$. These inequalities ensure the existence of a feasible experimentation policy where the scope of experimentation reaches its maximum of 2 with interior probability, while ensuring no other feasible policy yields a higher joint surplus. We list all the inequalities and comment on each one separately below.

$$
\begin{aligned}
2c &< \frac{\delta}{1-\delta}(\bar{v}-2c)+\mathcal{C}^0(0) && (36)\\
c &< \frac{\delta}{1-\delta}(\underline{v}-2c) && (37)\\
2c &> \frac{\delta}{1-\delta}(\underline{v}-2c)+\mathcal{C}^0(0) && (38)\\
2c &> \mathcal{C}^0(0)+\frac{\delta}{1-\delta}\Big(\Pr(\bar{v})\bar{v}+(1-\Pr(\bar{v}))\underline{v}-2c\Big) && (39)\\
\frac{\underline{v}-2c}{1-\delta} &> v := \mathbb{E}(v_p-2c)+\frac{\delta}{1-\delta}\Big(\Pr(\bar{v})(\bar{v}-2c)+\Pr(\underline{v})(\underline{v}-2c)\Big) && (40)\\
&\quad +\Big(\Pr(\bar{v})+\Pr(\underline{v})\Big)\frac{\delta}{1-\delta}(\underline{v}-2c)+\Big(1-\Pr(\bar{v})+\Pr(\underline{v})\Big)\delta v\\
c &\le \mathcal{C}^0(0)+\frac{\delta \Pr(\bar{v})\mathcal{C}^0(0)}{1-\delta(1-\Pr(\bar{v}))} && (41)
\end{aligned}
$$

Inequality (36) implies that $\{\bar{v},0\}$ satisfy Equation (5), where $\mathcal{C}^0(0)=B^0(0)-\mathbb{E}(v_p-2c)$ (i.e., the expected continuation value to player 0 when exploring in period 1. Inequality (37) ensures that the players are able to exploit a project worth $\underline{v}$ in isolation. Inequality (38) ensures that $|\mathbf{P}^t|<2$ while exploiting the project worth $\underline{v}$ (when the best project found so far on the other domain has value 0). This inequality uses $\mathcal{C}^0(0,\text{explore})$ as an upper-bound. These statements imply that if the players ever reach a point with a project worth $\underline{v}$, they either exploit the project, explore a project on the other domain while maintaining a scope of 1, or conduct 2 explorations. Inequality (39) ensures that conducting two explorations is not feasible because the upper-bounds associated with the continuation value for the new domain and the domain with a project with value $\underline{v}$ is provided by the first-best policy. Next, Inequality (40) ensures that the players prefer to exploit the project worth $\underline{v}$ as opposed to exploring the domain where the best project is worth 0 until Equation (5) holds and then subsequently implementing the first-best policy. These constraints imply that $|\mathbf{P}^t|=1$ if the best projects are worth $\underline{v},0$. Finally, Inequality (41) ensures that this experimentation policy is feasible. One can check that these constraints, along with (i) $\mathbb{E}\left(v_p\right)\ge 2c$ and (ii)

$v^0 \leq \underline{v}$, hold jointly.[25]

Statement 2: We consider $m = 3$ and a trinary support distribution, $v_p \in \{0, \underline{v}, \bar{v}\}$, where $0 < \underline{v} < \bar{v}$.[26] Throughout, let $\tilde{c} := c - (1-\alpha)\mathbb{E}(v_p) > 0$. Let $\bar{p}, \underline{p}$ correspond to the probability that $v_p = \bar{v}$, $\underline{v}$, respectively. For the subsequent argument, consider $\bar{p}$ to be arbitrarily small, in a sense we will make precise below. Suppose

$$2c = \frac{\delta}{1-\delta}(\bar{v} + \underline{v} - 4c), \tag{42}$$

implying that the players can jointly exploit projects worth $\bar{v}$ and $\underline{v}$, but could not permanently exploit a project worth $\bar{v}$ and two projects worth $\underline{v}$. As a result, for any $\tilde{c} > 0$, when $\hat{\mathbf{v}} = \bar{v}, \underline{v}, 0$, the players either (i) permanently exploit the two projects and conduct no additional explorations or (ii) implement a policy involving exploration whose associated surplus is bounded below that of exploiting the project with value $\bar{v}$ and exploring in the other two domains until finding a project with value $\bar{v}$ and subsequently implementing the first best. As the payoff bound of (ii) tends to zero as $\bar{p}$ goes to zero, there exists $\bar{p}^* > 0$ such that for $\bar{p} < \bar{p}^*$, the players choose (i). Hence, it suffices to show that Equation (42), $\bar{p} < \bar{p}^*$, and $|P^1| = m$ may jointly hold. Because the first two of these three conditions are independent of α, they are also independent of $\tilde{c}$. Therefore, holding fixed all remaining parameter values, the continuation value at date one of conducting three explorations is bounded below by $\Pr(v_p = \bar{v})\frac{\delta}{1-\delta}(\bar{v} - 2c)$. As a result, if $\tilde{c}$ is sufficiently small, the players' initial scope will be maximal, thereby completing the proof. □

References

Abreu, Dilip, "Extremal equilibria of oligopolistic supergames," *Journal of Economic Theory*, 1986, *39* (1), 191–225.

[25] The code can be provided upon request.

[26] Unlike Proposition 5, considering a trinary support distribution and $m = 2$ fails.

Acharya, Avidit and Juan Ortner, "Paths to the Frontier," *American Economic Journal: Microeconomics*, 2022, *14* (1), 39–69.

Albrecht, James, Axel Anderson, and Susan Vroman, "Search by committee," *Journal of Economic Theory*, 2010, *145* (4), 1386–1407. Search Theory and Applications.

Andrews, Isaiah and Daniel Barron, "The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents," *American Economic Review*, September 2016, *106* (9), 2742–59.

Anesi, Vincent and T. Renee Bowen, "Policy Experimentation in Committees: A Case against Veto Rights under Redistributive Constraints," *American Economic Journal: Microeconomics*, August 2021, *13* (3), 124–62.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *The Quarterly Journal of Economics*, 03 2017, *132* (3), 1101–1164.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, 1994, *109* (4), 1125–1156.

_ , _ , and _ , "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 2002, *117* (1), 39–84.

Benavides, Luis, Versa De Eskinazis, and Daniel Swan, "Six steps to successful supply chain collaboration," *CSCMP's Supply Chain Quarterly*, 2012, *2012* (2), 23–27.

Bergemann, Dirk and Juuso Välimäki, "Stationary multi-choice bandit problems," *Journal of Economic dynamics and Control*, 2001, *25* (10), 1585–1594.

_ **and Juuso Välimäki**, "Bandit Problems," in Steven Durlauf and Larry Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., Macmillan Press, 2008.

Bernard, Andrew B. and Andreas Moxnes, "Networks and Trade," *Annual Review of Economics*, 2018, *10* (1), 65–85.

Bernheim, B. Douglas and Michael D. Whinston, "Multimarket contact and collusive behavior," *The RAND Journal of Economics*, 1990, pp. 1–26.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," *The Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

_ **, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, " Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 11 2013, *128* (1), 1–51.

Board, Simon, "Relational Contracts and the Value of Loyalty," *American Economic Review*, December 2011, *101* (7), 3349–67.

Bolton, Patrick and Christopher Harris, "Strategic experimentation," *Econometrica*, 1999, *67* (2), 349–374.

Bull, Clive, "The Existence of Self-Enforcing Implicit Contracts," *The Quarterly Journal of Economics*, 1987, *102* (1), 147–159.

Callander, Steven, "Searching and Learning by Trial and Error," *American Economic Review*, October 2011, *101* (6), 2277–2308.

_ **and Bård Harstad**, "Experimentation in Federal Systems *," *The Quarterly Journal of Economics*, 04 2015, *130* (2), 951–1002.

— **and Niko Matouschek**, "The Risk of Failure: Trial and Error Learning and Long-Run Performance," *American Economic Journal: Microeconomics*, February 2019, *11* (1), 44–78.

Chan, Jimmy, Alessandro Lizzeri, Wing Suen, and Leeat Yariv, "Deliberating collective decisions," *The Review of Economic Studies*, 2018, *85* (2), 929–963.

Chassang, Sylvain, "Building routines: Learning, cooperation, and the dynamics of incomplete relational contracts," *American Economic Review*, 2010, *100* (1), 448–65.

Cisco Systems Inc., "Ericsson and Cisco Partner to Create the Networks of the Future," November 2015. Press release.

Doney, Patricia M. and Joseph P. Cannon, "An Examination of the Nature of Trust in Buyer-Seller Relationships," *Journal of Marketing*, 1997, *61* (2), 35–51.

Dwyer, F. Robert, Paul H. Schurr, and Sejo Oh, "Developing Buyer-Seller Relationships," *Journal of Marketing*, 1987, *51* (2), 11–27.

Ely, Jeffrey C and Juuso Välimäki, "Bad reputation," *The Quarterly Journal of Economics*, 2003, *118* (3), 785–814.

Garfagnini, Umberto and Bruno Strulovici, "Social Experimentation with Interdependent and Expanding Technologies," *The Review of Economic Studies*, 02 2016, *83* (4), 1579–1613.

Gibbons, Robert and Rebecca Henderson, "What Do Managers Do? Exploring Persistent Performance Differences Among Seemingly Similar Enterprises," *Handbook of Organizational Economics*, 2013, pp. 342–372.

Gonsalves, Antone, "Ericsson–Cisco Partnership Activity Expected to Slow Until 2020," July 2017.

Grossman, Sanford J. and Oliver D. Hart, “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 1986, *94* (4), 691–719.

Halac, Marina, “Relationship building: conflict and project choice over time,” *The Journal of Law, Economics, & Organization*, 2014, *30* (4), 683–708.

_ **, Navin Kartik, and Qingmin Liu**, “Optimal contracts for experimentation,” *The Review of Economic Studies*, 2016, *83* (3), 1040–1091.

Hart, Oliver and John Moore, “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, 1990, *98* (6), 1119–1158.

Hörner, Johannes, Nicolas Klein, and Sven Rady, “Overcoming Free-Riding in Bandit Games,” *The Review of Economic Studies*, 11 2022, *89* (4), 1948 – 1992.

Ide, Enrique, “Dual Moral Hazard and The Tyranny of Success,” *American Economic Journal: Microeconomics*, 2024.

Jovanovic, Boyan and Rafael Rob, “Long Waves and Short Waves: Growth Through Intensive and Extensive Search,” *Econometrica*, 1990, *58* (6), 1391–1409.

Keller, Godfrey, Sven Rady, and Martin Cripps, “Strategic Experimentation with Exponential Bandits,” *Econometrica*, 2005, *73* (1), 39–68.

Krieger, Joshua Lev, Christopher Stanton, and James Barnett, “Forecasting ClimaCell,” Dec 2019.

Levin, Jonathan, “Multilateral contracting and the employment relationship,” *The Quarterly Journal of Economics*, 2002, *117* (3), 1075–1103.

_ , “Relational Incentive Contracts,” *American Economic Review*, June 2003, *93* (3), 835–857.

Liker, Jeffrey K and Thomas Y Choi, "Building deep supplier relationships," *Harvard business review*, 2004, *82* (12), 104–113.

Lippman, Steven A. and John J. McCall, "The Economics Of Job Seach: A Survey," *Economic Inquiry*, 1976, *14* (2), 155–189.

Liu, Qingmin and Yu Fu Wong, "Strategic Exploration: Pre-emption and Prioritization," *Review of Economic Studies*, 2023, p. rdad084.

Lizzeri, Alessandro, Eran Shmaya, and Leeat Yariv, "Disentangling Exploration from Exploitation," Technical Report, National Bureau of Economic Research 2024.

Macchiavello, Rocco and Ameet Morjaria, "The value of relationships: evidence from a supply shock to Kenyan rose exports," *American Economic Review*, 2015, *105* (9), 2911–2945.

Macleod, W. Bentley and James Malcomson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 1989, *57* (2), 447–80.

Marek, Sue, "Cisco Ericsson Partnership Will Miss $1B Revenue Target for 2018," November 2017.

Matsumoto, Craig, "Cisco & Ericsson Hone Their Partnership," June 2017.

Monarch, Ryan, "It's Not You, It's Me: Prices, Quality, and Switching in U.S.-China Trade Relationships," *The Review of Economics and Statistics*, 09 2022, *104* (5), 909–928.

_ **and Tim Schmidt-Eisenlohr**, "Longevity and the value of trade relationships," *Journal of International Economics*, 2023, *145*, 103842.

Reshidi, Pëllumb, Alessandro Lizzeri, Leeat Yariv, Jimmy Chan, and Wing Suen, "Sequential Sampling by Individuals and Groups: An Experimental Study," American Economic Review: Insights 2025. Forthcoming.

Robbins, Herbert, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, September 1952, *58* (5), 527–535.

Spolaore, Enrico, "The Political Economy of European Integration," NBER Working Paper 21250, National Bureau of Economic Research, June 2015.

Strulovici, Bruno, "Learning while voting: Determinants of collective experimentation," *Econometrica*, 2010, *78* (3), 933–971.

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, June 2011, *49* (2), 326–65.

Tirole, Jean, *The Theory of Industrial Organization*, Cambridge, MA: MIT Press, 1988.

Urgun, Can, "Restless Contracting," 2021. Working Paper.

Vanpoucke, Evelyne, Ann Vereecke, and Kenneth K. Boyer, "Triggers and patterns of integration initiatives in successful buyer–supplier relationships," *Journal of Operations Management*, 2014, *32* (1), 15–33.

Venables, Sarah, "Essays on organizational economics and the heterogeneity of firm performance." Ph.d. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA 2013.

Wang, Shaoda and David Y. Yang, "Policy Experimentation in China: The Political Economy of Policy Learning," *Journal of Political Economy*, forthcoming.

Watson, Joel, "Starting small and renegotiation," *Journal of Economic Theory*, 1999, *85* (1), 52–90.

_ , "Starting Small and Commitment," *Games and Economic Behavior*, 2002, *38* (1), 176–199.

Weitzman, Martin, "Optimal Search for the Best Alternative," *Econometrica*, 1979, *47* (3), 641–654.

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