Multi-Project Collaborations

Charles Angelucci and Roi Orzach*

August 6, 2024

Abstract

We analyze collaborative experimentation across multiple independent domains, each containing 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. We characterize the optimal experimentation policy for two domains, which exhibits common features of collaborative experimentation: lengthy exploration, gradual scope expansion, permanently bounded scope, intermittent domain exploration, and project revival. We connect these findings to research on buyer-supplier dynamics and persistent productivity differences.

Keywords: Strategic Experimentation, Relational Contracting, Gradualism
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, MIT, 50 Memorial Drive, Cambridge MA 02142 (e-mail: orzach@mit.edu).
We thank Abhijit Banerjee, 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, 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 contexts, actors innovate collaboratively across multiple domains at once. In buyer-supplier relationships, companies co-innovate across multiple product lines or geographies. In the pharmaceutical sector, an R&D alliance may combine resources across vaccine development and protein targeting. Inside firms, continuous improvement methods involve managers and workers collaborating to identify and implement improvements across various dimensions of the production process.

The success of these collaborations relies on maintaining aligned interests, 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 by withdrawing their cooperation—creates interdependencies across domains. For instance, in a collaboration involving two domains, a breakthrough in one domain will increase the parties' perceived value of the collaboration, mitigating opportunism in the other domain. In innovation-intensive settings, parties must therefore approach their joint experimentation in each domain of cooperation by balancing the domain-specific outcomes with the broader implications for the overall collaboration. This paper investigates how these cross-domain interdependencies influence the dynamics and outcomes of joint experimentation. Furthermore, it relates the main theoretical findings to the existing applied literature on buyer-supplier dynamics and persistent productivity differences across firms.

We develop a model of multi-domain collaborative innovation by extending a canonical experimentation framework. Time is discrete with an infinite horizon, and the number of domains is exogenous. Each domain contains infinitely many ex ante identical projects on which the players can cooperate, and the domains are technologically independent. Cooperation on a project requires both players' participation. In each period and domain, players can choose to idle, explore a new project, or exploit a known one. 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 whether to exploit a known project or explore a new one for each domain—must be self-enforcing.

We focus on Subgame Perfect Equilibria (relational contracts) that maximize the players' discounted cumulative joint payoffs (their "surplus"). As a starting point, Proposition 1 examines our first benchmark, the single-player scenario, providing a straightforward solution in which the optimal experimentation policy treats each domain 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. Henceforth, we call this optimal policy for the single-player scenario the "first-best experimentation policy." Two players collaborating in experimentation would optimally adopt this policy if all projects benefited each player equally.

In the main analysis, we introduce asymmetric project benefits by assuming each project exclusively benefits one player. The beneficiary's identity, initially unknown, is revealed when players first cooperate on the project and is independently and identically distributed across projects. The presence of asymmetric project benefits is the key friction potentially impeding the implementation of the first-best experimentation policy, as this policy requires promises of transfers between players to ensure cooperation. However, these promises may lack credibility if the players making the transfers have insufficient continuation value in the collaboration. As mentioned above, since any player's continuation value from the collaboration corresponds to the sum of the continuation values associated with each domain of cooperation, the players' experimentation choices in one domain will affect all domains. For example, the decision to exploit a given project in one domain becomes directly linked to the progress made or anticipated in other domains. These cross-domain dynamics make an analytical characterization of the optimal experimentation policy challenging.

We make progress by combining standard dynamic programming tools with the insight from Levin (2002, 2003) that, with transferable utility, a single constraint can capture all deviation temptations across players, domains, and outcomes from project explorations conducted in the current period (Proposition 2). This approach simplifies equilibrium characterization, enabling us to identify key properties of the optimal experimentation policy. Proposition 3 establishes a necessary and sufficient condition for implementing the first-best experimentation policy despite asymmetric project benefits, based on the values of the most valuable projects identified to date in each domain. This condition requires the joint value of these projects to be high enough to ensure the collaboration's continuation value supports implementing the first-best policy. For low discount factors, this condition binds, implying that, in expectation, players transition to permanently exploiting the most valuable projects in each domain later than if they could implement the first-best from the relationship's start (Corollary 1). In some cases, this transition may never occur, as discussed below. Moreover, this condition enables us to fully characterize the optimal experimentation policy in our second benchmark analysis: the single-domain case. In this simplified setting, exploration continues until a project's value exceeds a time-invariant threshold, higher than that in the single-agent benchmark, after which permanent exploitation of the project becomes optimal (Corollary 2).

Next, we analyze the properties of the players' optimal experimentation policy when they are unable to implement the first-best policy in the current period. For tractability we focus on the case of two domains. Proposition 4 shows that the players may find it optimal to gradually expand the scope of their collaboration by exploring a new domain only after identifying a sufficiently valuable project in a first domain. This approach is optimal when exploring all domains simultaneously from the outset is not feasible, due to a low initial continuation value of the collaboration. A gradual policy can be easier to implement because it helps the players to collaboratively experiment in each of the two domains. The exploration of the first domain is easier because it is motivated not only by its potential outcome within that domain but also by the added prospect of exploring the second domain upon identifying a sufficiently valuable project in the first. The exploration of the second domain is easier because its exploration begins once a sufficiently valuable project in the first domain has increased the players' overall continuation value. Finally, while a gradual experimentation policy requires a positive probability of exploring an additional domain to successfully initiate the collaboration, Proposition 5 shows that this expansion may never take place. When scope expansions are not guaranteed to occur, the long-term scope of the collaborative relationship is determined by the outcomes of the early project explorations.

Having examined the implications of cross-domain interdependencies on the dynamics of the scope of players' collaboration, we conclude the main analysis by investigating their consequences for the players' exploration and exploitation decisions within their active domains of cooperation. Unlike the first-best experimentation policy, which features a time-invariant exploitation threshold, the threshold for project exploitation is dynamic when the players' continuation value is initially low in multi-domain experimentation. In Proposition 6, we show that this dynamic threshold can give rise to two notable behaviors. First, the players may exploit a project temporarily, meaning that they may choose to stop exploring a domain for an extended period and later decide to resume exploration. Second, the players may choose to stop exploring a domain, but instead of exploiting the most recently explored project, they may opt to exploit a project they had previously explored but not exploited. Both behaviors are driven by the evolving aggregate value of the collaboration.

In Section 5, we discuss extensions of the model included in the Online Appendix, in which the domains of cooperation are asymmetric or exhibit technological interdependencies. We find that these features increase the importance of gradual experimentation policies. We also examine how the potential scope of a collaboration influences its feasibility and profitability.

Section 6 connects our theoretical analysis with two distinct research areas: buyer-supplier dynamics and persistent productivity differences across firms. The buyer-supplier relationships literature stresses experimentation and credibility as critical factors for successful collaborations, and corroborates the prevalence of gradualism and strong path dependence. In addition, we argue that our framework provides novel 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. Section 5 discusses various model extensions. Section 6 examines the applied literature in light of our theoretical findings. Section 7 concludes the paper.

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).

For a review of applications within economics, see Bergemann and Välimäki (2008). Our work contributes to the subset of this literature that focuses on strategic interactions among players. Bolton and Harris (1999) and Keller et al. (2005) consider settings in which players independently pull arms and free-ride on each others' experimentation (see also Hörner et al., 2022, for more recent work on this topic). Bonatti and Hörner (2011) examine a scenario in which agents collaborate in an experimentation process involving private effort choices. Further, Liu and Wong (2023) consider an environment in which players compete with each other to explore alternatives. Our focus is on a setting where cooperation among players is essential for both the exploration and exploitation of projects, as individual experimentation is not feasible. In Strulovici (2010), players vote to choose between a safe arm and a risky one, with its asymmetric benefits revealed over time through experimentation. Anesi and Bowen (2021) analyze a similar setting, allowing for some redistribution of surplus among players. Further, Albrecht et al. (2010) examine a sequential search problem where a committee determines which project to permanently exploit. Chan et al. (2018) and Reshidi et al. (2024) contrast group and individual decision-making regarding experimentation, looking at the impact of static versus sequential information acquisition and of voting rules. In contrast to these papers, our setting allows for voluntary transfers among players and requires the combined efforts of all players for experimentation. Most significantly, the distinguishing feature of our framework is that players can simultaneously experiment across multiple domains.

Secondly, this work is related 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).[1] Building upon the work of Levin (2002, 2003), we leverage the key insight that, in the presence of monetary transfers, all constraints associated with players' temptations to deviate can be aggregated into a single constraint. Halac (2014) studies a repeated principal-agent setting in which the value of the players' relationship increases with the duration of the relationship. The players initially choose to cooperate on low-risk, low-return projects, and they switch to high-risk, high-return projects once their relationship has grown sufficiently valuable. In our setting, experimentation endogenously affects the value of the players' relationship,

[1]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.

creating a feedback effect: as players engage in experimentation, their relationship grows more valuable, which in turn facilitates more efficient experimentation. In contrast, Chassang (2010) analyzes a setting where increases in relationship value diminish the players' ability to experiment. In his model, the agent knows which arms are productive and which are not, while the principal, at the outset, cannot differentiate between the two. 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, the players are symmetrically informed about their environment, and the presence of transferable utility removes the need for inefficient on-path punishments. These two features lead to the positive feedback effect mentioned above.[2]

Finally, we add to the body of research on gradualism in collaborations. Watson (1999, 2002) examine a setting in which players are uncertain regarding their counterpart' intentions—to either collaborate genuinely or take advantage of the other. The players begin with low cooperation to mitigate the losses from defection. As the players become more optimistic, the collaboration grows. Collaborations involving trustworthy players achieve optimal cooperation, while those with untrustworthy players eventually fail. In our setting, the scope of the players' experimentation may increase incrementally, not due to screening intentions, but because building the continuation value needed for self-enforcing experimentation takes time.[3] As a result, the two settings make opposite predictions regarding the impact of the discount factor on gradualism. In our setting, a higher discount factor reduces the need for gradualism, whereas in the frameworks analyzed by Watson (1999, 2002), and the dynamic

[2]Introducing transferable utility within Chassang (2010), where information asymmetry plays a central role, would make the value of the players' relationship constant on path. 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).

[3]Gradualism also arises in Ghosh and Ray (1996) and Kranton (1996), where players are randomly matched and can exit relationships at any time, with new partners possessing only limited information about the player's past history. In the context of corporate finance, investment levels can increase over time as borrowers gradually build collateral (see Tirole, 2006, Chapter 4 and references therein). In our scenario, the continuation value of the relationship functions like a collateral. However, this collateral is pledged by both players involved, rather than a single player. Finally, gradualism exists in other dynamic screening settings, such as in Ely and Välimäki (2003), where the players' relationship gradually becomes more efficient as their beliefs become more optimistic.

screening literature more broadly, a higher discount factor increases gradualism because separation is harder to achieve.

2 The Setup

Two players, such as a buyer and a supplier or two firms in an R&D alliance, 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 $v_p \in \mathbf{R}$, 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.[4] The identity of any project's beneficiary is, however, initially unknown and 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 $1 - \alpha$.

A project p 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. There are no intertemporal restrictions on project availability.

We make the following assumptions on the distribution of project values. First, we assume that the distribution of v_p has a convex support. Second, we assume $\mathbb{E}(v_p) \geq 2c$, which is a sufficient condition to ensure that exploration increases joint surplus.[5]

[4] Our results hold for less skewed benefit distributions, provided one player's valuation exceeds the cost c while the other's falls below it for each project.

[5] Relaxing this assumption could make no experimentation optimal in the first-best for low discount factors, unnecessarily complicating our analysis of the second-best policy where the discount

These assumptions simplify the proofs and statements of the results but are not crucial. We further assume that with positive probability, $v_p \geq \tilde{v}(\delta) := c\,(1+\delta)\,/\delta$. As Proposition 2 will show, without project values exceeding this threshold, players will never cooperate in exploiting projects.

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 $w^t_{i,-i} \in \mathbf{R}^+$ 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 $b^t_{i,-i} \in \mathbf{R}^+$ denotes such a transfer from player i to player $-i$.[6] 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 $\mathbf{x}^t$. Such a relational contract is self-enforcing if it describes a Subgame Perfect Equilibrium of the repeated game. Within the class of Subgame Perfect Equilibria, we analyze pure-strategy equilibria which maximize the players' joint surplus. Pure strategy equilibria are optimal because (i) mixing on transfers in-

factor is key.

[6] 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.

creases the maximal transfers players promise each other 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).[7]

3 First-Best Experimentation

We characterize the optimal experimentation policy for a benchmark where a single decision-maker, "player 0," maximizes the sum of both players' payoffs. This optimal experimentation policy is identical to the one we would obtain if we modified the model described in Section 2 such that projects always equally benefit both players. The proofs for the following proposition, as well as those for all subsequent statements not included in the main text, can be found in the Appendix.

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; If no previously-explored project has a value exceeding $v^0(\delta)$, explore a new project. The threshold $v^0(\delta)$ is increasing in δ.

Player 0 treats each domain separately and identically, given the additive separability of payoffs across projects and domains, as well as the ex ante identical nature of domains. The threshold v^0 arises from player 0's decision in each domain to either exploit the best project found thus far or explore a new project in search of a superior one. Furthermore, exploitation is permanent because player 0 does not acquire new information when exploiting a project. Likewise, given the infinite supply of ex ante identical projects in every domain, player 0 never chooses to exploit a project it chose not to exploit in the past. Finally, as the discount factor increases, the value of exploration rises because any superior project identified can be exploited across all future periods, explaining the comparative static result for v^0.

[7] Alternatively, post-deviation, 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.

In summary, the first-best policy maximizes experimentation scope, with exploration/exploitation decisions in each domain dictated by an independent, identical, and time-invariant threshold. We now analyze the model from Section 2, where these features may not always hold.

4 Main Analysis

This section analyzes the model described in Section 2. 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 experimentation policy stated in Proposition 1. In Section 4.3, we characterize key properties of the players' optimal experimentation policy when they are unable to implement the first-best policy. For tractability, Section 4.3 focuses on the case where $m = 2$. Finally, in Section 4.4, we analyze a concrete example to illustrate some of the key model dynamics.

4.1 Optimal Experimentation Policies: Implementability

In our setting, surplus-maximizing relational contracts depend on the players' beliefs about the projects. We denote the beliefs at the beginning of period t 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 players' beliefs μ^t. Furthermore, the continuation equilibria prescribed by such surplus-maximizing relational contracts are also 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 <u>optimal</u>. 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.

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 $\mu^t\left(h_1^t\right) = \mu^{t'}\left(h_2^{t'}\right)$, 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 $h^t \in \mathcal{H}^t$:*

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

  where $\mathcal{C}\left(\mu^t\right)$ ("the continuation value") is the expected net present value of the players' joint surplus starting in $t+1$ given $\hat{\mathbf{P}}(\cdot)$ and μ^t.

The proof of this proposition, provided in the Online Appendix, 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, Inequality (2) states that for an optimal relational contract to implement an experimentation policy everywhere on path, the continuation value must exceed the total reneging temptation across players and projects in all periods and histories. 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 induced by the policy. This simplification arises because all transfers cancel out in both the joint surplus expression and the right-hand side of (2). Notably, Inequality (2) implies that players might not treat domains independently, as its right-hand side aggregates continuation values across all domains of cooperation. The consequences of this "relational" interdependence across domains on the players' optimal experimentation policy will be our focus hereafter.

4.2 Implementability of First-Best Experimentation

In this section, we provide necessary and sufficient conditions on the beliefs μ^t under which the players can implement the first-best experimentation policy described in Proposition 1 in all periods $t' \geq t$. We refer to this outcome as "implementing the first-best experimentation policy." As we will show, in equilibrium, it may happen that there exists 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.

Intuitively, the optimal relational contract in any given period depends solely on the values of the most valuable projects identified in each of the m domains of cooperation, which we denote as $\hat{v}_1, \ldots, \hat{v}_m$.[8] Less valuable projects will never be exploited. As a result, keeping track of $\hat{v}_1, \ldots, \hat{v}_m$ is sufficient to keep track of the players' beliefs about the projects. Moreover, Inequality (2) implies that the threshold $\tilde{v}$, defined in Section 2, represents 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, which entails exploiting a project if and only if its value is at least v^0.

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:*

$$h(\hat{v}_1, \ldots, \hat{v}_m) := \frac{1}{m} \sum_{j=1}^{m} \max\{\hat{v}_j, v^0\} \geq \tilde{v} := c\frac{1+\delta}{\delta}. \tag{3}$$

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 (3) is satisfied, the continuation value of the relationship is sufficiently high to enable the implementation of the first-best experimentation policy.

[8]Formally, if no projects have been explored in domain j, we set $\hat{v}_j = 0$.

The condition requires that the average across domains of the maximum between the value of the most valuable project found in each domain and the threshold v^0 must exceed the threshold $\tilde{v}$. The function $h(\hat{v}_1, \ldots, \hat{v}_m)$ is not the arithmetic mean of the values $\hat{v}_1, \ldots, \hat{v}_m$ for two reasons: (i) under the first-best policy, players explore rather than exploit projects with values lower than v^0, and (ii) exploration contributes to the players' continuation value. Furthermore, the condition $v^0 \geq \tilde{v}$ is both necessary and sufficient for Inequality (3) to hold from period 1 onwards. The function $v^0(\delta) - \tilde{v}(\delta)$ exhibits a single-crossing property in δ, implying the existence of a threshold δ^0.[9] Players implement the first-best policy from period 1 if and only if $\delta \geq \delta^0$.

Next, Proposition 3 allows us to provide necessary and sufficient conditions under which the players cease all exploration and transition to exploiting the most valuable project discovered in each domain, given 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.

$$\{\hat{v}_j \geq v^0 \, \forall j \text{ and } \frac{\sum_j \hat{v}_j}{m} \geq \tilde{v}\} \iff \{\hat{v}_j \geq v^0 \, \forall j \text{ and } \frac{\sum_j \max\{\hat{v}_j, v^0\}}{m} \geq \tilde{v}\} \qquad \square$$

The conditions stated in Corollary 1 imply that, in expectation, the players attain 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 shown below, permanent exploitation across all domains of cooperation is not even guaranteed to occur.

We conclude by noting that the conditions listed in Corollary 1 fully characterize the players' optimal experimentation policy for the second natural benchmark case in our analysis: a single-domain collaboration. When there is only one domain, 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.

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

Corollary 2 (**Single-Domain Experimentation Benchmark**)
When $m = 1$, 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. Finally, there exists a threshold $\delta^* \geq 0$ such that the optimal relational contract is non-empty if and only if $\delta \geq \delta^*$.

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. We have also shown that, if δ is not sufficiently high, the players will initially be unable to implement the first-best policy. We now proceed to characterize key properties of the players' experimentation policy in the periods that precede an eventual transition to the first-best policy when collaboration spans multiple domains.

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, referring to this as their second-best experimentation policy. In each domain, players either (i) explore a new project, (ii) exploit the project with the highest known value $\hat{v}_j$, or (iii) remain idle. With m domains, there are thus 3^m possible action combinations each period. Since our proofs rely on analytical solutions, for tractability we focus on the case where $m = 2$ in this subsection. While the parity of m may influence certain aspects of the optimal experimentation policy (as discussed in Section 5.2), the core insights derived in this section apply to cases with larger values of m.

The players' exploration and exploitation decisions within their active domains of collaboration are inherently intertwined with their choices of which domains to engage in. However, to disentangle these dynamics and help intuition, we examine both aspects of the players' experimentation separately. In Section 4.3.1, we examine the dynamics of the scope of the players' experimentation, setting their exploration and exploitation decisions aside. Subsequently, in Section 4.3.2, we reverse the focus, exploring the dynamics of their exploration and exploitation decisions while leaving their scope decisions in the background.

4.3.1 The Dynamics of the Scope of Experimentation

Proposition 3 established a threshold δ^0 on the discount factor such that, when $\delta \geq \delta^0$, players implement the first-best policy from period 1, maintaining maximal scope throughout. We now show that, when $\delta < \delta^0$, players may instead expand the scope of collective experimentation gradually over time.

Proposition 4 (**Gradualism in the Scope of Experimentation**)
Fix $m = 2$. There exist two additional thresholds $\delta^* \leq \bar{\delta} \leq \delta^0$ on δ such that:

1. *If $\delta \geq \bar{\delta}$, any optimal relational contract is such that the scope of experimentation is always maximal on path (i.e., $|\mathbf{P}^t| = 2$ for all t).*

2. *If $\delta \in [\delta^*, \bar{\delta})$, any optimal relational contract is such that the scope of experimentation is initially limited on path (i.e., $|\mathbf{P}^1| = 1$), with scope increasing with strictly positive probability along the equilibrium path.*

3. *If $\delta < \delta^*$, the scope of experimentation is equal to zero in all periods (i.e., $|\mathbf{P}^t| = 0$ for all t).*

There exist distinct open sets of parameter values such that the interval $[\delta^*, \bar{\delta})$ is non-empty for one set and empty for another set.

When $\delta \geq \bar{\delta}$, the scope of experimentation is maximal from the start, though the first-best policy may be delayed. When $\delta < \delta^*$, no experimentation occurs at any time. If $\delta^* < \bar{\delta}$, for $\delta \in \left[\delta^*, \bar{\delta}\right)$, the scope of experimentation gradually increases over time. These thresholds exist because (i) it is optimal for the players to experiment in as many domains as possible, as soon as possible, within the constraints of Inequality (2), and (ii) because a higher δ relaxes these constraints.

We now explore why players might find it optimal to gradually expand their experimentation domains. This result seems counterintuitive for three reasons: (i) Exploration is beneficial even in the current period (as $\mathbb{E}(v_p) \geq 2c$ by assumption); (ii) Exploration provides valuable information, increasing the domain's continuation value; (iii) Due to (ii), exploration relaxes Inequality (2), potentially improving efficiency in other domains' decision-making. However, gradualism reduces the left-hand side of Inequality (2) early in the relationship, making this approach potentially optimal despite these considerations.

To better understand this trade-off, we compare three experimentation policies, focusing on the most stringent constraint imposed by Inequality (2) for each. This approach is sufficient because an experimentation policy is implementable if and only if this inequality holds throughout the equilibrium path. Also, we set $\alpha = 1$ for simplicity, making the left-hand side of Inequality (2) equal to player 2's total cost. Under the first policy ("Independent"), players treat both domains independently and identically. The most stringent constraint occurs in period 1, when players explore their first project in each domain:

$$2c \leq 2\,\mathcal{C}(\text{exploration}). \tag{4}$$

Under the second policy ("Maximal"), players immediately explore both domains, but the decision to exploit projects is interdependent across domains. The most stringent constraint still applies in period 1:

$$2c \leq \mathcal{C}(\text{two explorations}). \tag{5}$$

Under the third policy ("Gradual"), players start by exploring domain 1 and expand to domain 2 upon discovering a sufficiently valuable project. The most binding constraint corresponds to either period 1 (6) or the period of expansion (7):

$$c \leq \mathcal{C}(\text{exploration}) + \mathcal{C}(\text{delayed exploration}), \tag{6}$$

$$2c \leq \mathcal{C}(\text{exploitation}) + \mathcal{C}(\text{exploration}). \tag{7}$$

Inequality (4) is the most stringent constraint of the the four, reflecting the advantages players derive from pooling their two domains of experimentation.[10] It is harder to satisfy than Inequality (5) because, under the Maximal policy, players can use the discovery of a valuable project in one domain to make more efficient exploration and exploitation decisions in the other (e.g., by lowering the exploitation threshold). Moreover, Inequality (6) is easier to satisfy than (4) due to the positive continuation value stemming from the eventual exploration of the second domain. Finally, Inequality (7) is also easier to satisfy than (4) because $\mathcal{C}(\text{exploitation}) > \mathcal{C}(\text{exploration})$ when exploiting a domain 1 project that enables the exploration of domain 2.

[10] In fact, Inequality (4) corresponds to the implementation constraint when the number of domains equals 1. See Section 5.2 for a discussion of how scope affects the players' experimentation.

We can now more readily examine the conditions determining whether the interval $[\delta^*, \bar{\delta})$ is empty or non-empty. Inequality (7) is always easier to satisfy than Inequality (5), for reasons similar to why it is easier to satisfy than Inequality (4). Therefore, we focus on comparing Inequalities (5) and (6). Intuitively, the Maximal policy is optimal whenever feasible and it can be feasible for lower discount factors than the Gradual policy. The drawback of the Gradual policy is that it shifts surplus into the future, potentially reducing both profitability and feasibility. Examining Inequalities (5)-(6), this occurs whenever $\mathcal{C}$(delayed exploration) is low, such as when the discovery of a valuable domain 1 project, which enables the exploration of domain 2, is expected to take a long time. In these instances, the interval $[\delta^*, \bar{\delta})$ is empty.

Conversely, the interval $[\delta^*, \bar{\delta})$ is non-empty and gradual scope expansion is optimal for intermediate values of the discount factor if $\mathcal{C}$(delayed exploration) is not significantly lower than $\mathcal{C}$(exploration), and/or if the benefit of simultaneous exploration in two domains with interdependent exploration/exploitation decisions is not substantially greater than making independent decisions (i.e., $\frac{1}{2}\mathcal{C}$(two explorations) $\approx$ $\mathcal{C}$(exploration)). This scenario may occur when the continuation value of the players' collaboration approaches that needed for the implementation of the first-best experimentation policy, such that discovering a valuable project in one domain minimally affects the exploitation threshold in the other. This would happen, for example, if the distribution of project valuations, v_p, assigns little density to realizations between $2c$ and $2\tilde{v} - v^0$ (the minimum value for $\hat{v}_j$ under which (3) holds regardless of $\hat{v}_{-j}$).[11] As this intermediate range's density approaches zero, the discovery of any project worth exploiting enables first-best experimentation. Consequently, the benefit of interdependent exploration/exploitation decisions across domains diminishes, making Inequality (6) easier to satisfy than Inequality (5).

While we set $\alpha = 1$ to simplify Inequality (2) and illustrate why gradualism may be optimal, the same basic intuitions hold when $\alpha < 1$. Gradualism can be optimal for intermediate values of the discount factor because it reduces the present cost of experimentation while only delaying the exploitation of valuable projects in other domains. Although this logic requires a non-zero probability of future scope expansion, we now show that an optimal gradual experimentation policy does not necessarily imply expanding the scope of collaboration with certainty.

[11] For this reason, the interval $[\delta^*(\alpha), \bar{\delta}(\alpha))$ is always non-empty for binary distributions.

Proposition 5 (**Bounded Scope**) *Fix $m = 2$. There exists an open set of parameter values such that any optimal relational contract is non-empty and the broadest scope of the players' experimentation achieved on the equilibrium path need not reach its maximum with probability one.*

Consider a discount factor such that the scope of experimentation increases gradually. Players begin by exploring domain 1. They may find a project valuable enough for exploitation and domain 2 exploration, or one supporting only permanent exploitation. In the latter case, players might choose limited scope if the chances of finding a significantly better project are low. The seeming paradox—inability to explore domain 2 after a valuable domain 1 discovery, despite initial domain 1 exploration—is resolved by recognizing that initial exploration was partly motivated by potential scope expansion.

In this subsection, we have shown that, for intermediate values of the discount factor, the optimal second-best experimentation policy may involve a gradual increase in the number of domains in which players collaboratively experiment. While the possibility of scope expansion must have a positive probability for collective experimentation to begin, it is not guaranteed, and players may ultimately remain permanently idle in a domain. We now shift our focus to analyzing the dynamics governing players' exploration and exploitation decisions.

4.3.2 The Dynamics of Exploration-Exploitation Decisions

Proposition 1 described the single-player benchmark's optimal experimentation policy (the first-best policy). In this policy, each domain is treated independently and identically, and the threshold for project exploitation is time-invariant. This latter property implied that project exploitation and the decision to not exploit a project were permanent.

In contrast, for collaborative experimentation, we have shown in previous sections that players aggregate incentives across all their cooperative domains. This pooling generates relational interdependencies, resulting in domains being treated neither identically nor independently. In this section, we establish that this observation implies that the criterion used to determine project exploitation is dynamic. In particular, we show that the players may exploit a project temporarily, and further, that they may recall a project they previously chose not to exploit.

Proposition 6 (**Temporary Exploitation and Recall of Projects**)
The players' optimal experimentation policy may involve temporary exploitation and recall of projects. Specifically, for an open set of parameter values and with strictly positive probability:

1. *The players may choose to exploit a project in period t, but later decide not to exploit the same project in some period $t' > t$.*
2. *The players may choose not to exploit a project in period t, but later decide to exploit the same project in some period $t' > t$.*

The behaviors described in the proposition do not occur for all parameter values. They never occur if $\delta \geq \delta^0$ and the players implement the first-best experimentation policy from the start. Even if $\delta < \delta^0$, these behaviors are possible but not guaranteed. For example, they will not occur if the players immediately identify two projects worth permanent exploitation in the initial period, as there would be no incentive to deviate from exploiting these projects in the future.

The intuition behind (1) can be understood by considering the following scenario. Suppose the values of the best projects in domains 1 and 2 satisfy $\hat{v}_1 \geq \hat{v}_2$. Further, assume that both values are sufficiently large for the players' relationship scope to be maximal, but not large enough to enable them to implement the first-best policy. If $\hat{v}_1$ is particularly high, the players will choose to exploit the project in domain 1 and explore in domain 2. Now, imagine that the exploration in domain 2 uncovers a project with a value slightly higher than $\hat{v}_1$. In this case, the players find themselves in a situation similar to the previous period, but with the roles of the domains reversed. They will now choose to exploit the newly discovered project in domain 2 and explore in domain 1.

To understand the intuition behind (2), consider a scenario where the discount factor δ is sufficiently small, preventing the exploitation of projects with values only slightly above the exploitation threshold v^0. Suppose the players' scope of experimentation is maximal, which for instance occurs when α equals 1/2. If the explorations in period 1 yield two projects with values marginally higher than v^0, the players are forced to explore two new projects in the next period. However, if one of these newly explored projects achieves a significantly high value, it can substantially increase the continuation value of the players' relationship. This increased continuation value may enable the players to implement the first-best experimentation policy. In such a case, the players may find it optimal to revert to exploiting one of the two projects from period 1, even though they previously chose to explore new projects.

Behaviors such as the temporary exploitation of projects or the recall of past projects are common in experimentation settings. These behaviors can arise due to various factors, including the presence of a finite number of projects to explore or the fact that project characteristics may not be fully revealed immediately. Our analysis has shown that strategic interactions alone can also drive these behaviors.

4.4 Multi-Project Collaborations: A Graphical Illustration

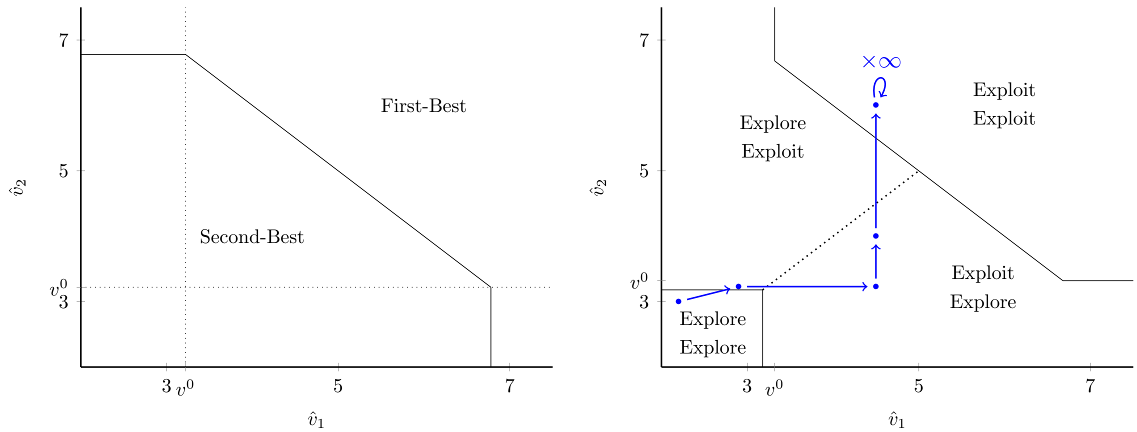
We analyze an example with specific parameter values. We set $c = 1$ and $\delta = 1/3$. Furthermore, we consider a symmetric relationship by setting $\alpha = 1/2$. The players can cooperate in two domains ($m = 2$). Finally, the project values v_p are drawn from a shifted exponential distribution with 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 inactivity. 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 (3) 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. In contrast, 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 $v_p \sim 1 + \mathrm{Exp}(1 \,/\, 2)$. $\hat{v}_1$ 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.

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. 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.[12]

[12] The threshold v' presented in the figure 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.

Figure 1b also presents a sample path illustrating the evolution of realized project values over time, depicted in blue. In the early phase where the players are exploring two projects simultaneously, 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 of the relationship 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 6.

5 Further Results

This section extends our analysis in two directions. First, we explore several simple extensions in which the domains of cooperation are not identical or independent. Subsequently, we examine how the potential scope of a collaboration influences its feasibility and profitability.

5.1 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 natural 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 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 the 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 made the assumption 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. The other domain follows the main analysis, where project benefits accrue exclusively to one player. We show that gradualism is always optimal 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 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, a gradual approach is strictly optimal for intermediate discount factors. These findings suggest gradualism is more likely to be optimal in R&D environments with stronger cross-domain knowledge spillovers.

5.2 The Benefits of Scope

The maximum potential scope of collaboration, denoted by m in our analysis, can vary significantly depending on the specific 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 examine how variations in m influence the profitability and sustainability of collaborations.

Let $\tilde{\pi}(m) := \pi(m) \,/\, m$ denote the average joint surplus per domain of the collaboration. Similarly, let $\delta^*(m)$ represent the minimum discount factor for which the optimal relational contract is non-empty. For a scaling factor $k \geq 1$, the following weak inequalities hold: $\tilde{\pi}(m \cdot k) \geq \tilde{\pi}(m)$ and $\delta^*(m \cdot k) \leq \delta^*(m)$. The intuition behind these inequalities is that players can always engage in k independent and concurrent experimentation policies, each replicating the optimal collaboration with m domains. Consequently, scaling up the maximum potential scope cannot decrease $\tilde{\pi}$ or increase δ^* (c.f., Bernheim and Whinston, 1990).[13]

We can provide necessary and sufficient conditions for these inequalities to hold strictly. Specifically, $0 < \delta^*(m \cdot k) < \delta^*(m)$ for $k > 1$ if $(1-\alpha)\mathbb{E}(v_p) < c$ and otherwise $\delta^*(m \cdot k) = 0$ regardless of k. The condition $(1-\alpha)\mathbb{E}(v_p) < c$ implies that exploration is not an equilibrium of the stage game, as it is not in player 2's interest. When exploration is not an equilibrium of the stage game, 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. Instead, 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 first-best experimentation policy is not implementable at date zero. If the first-best policy is implementable, increasing the collaboration's maximum scope does not increase the average joint surplus.

[13] Also, while $\pi(m)$ is monotonically increasing in m, $\tilde{\pi}(m)$ may not be. To see this, suppose that v_p belongs to a three-point support (low, medium, and high). Suppose further that a high-valued project and a single medium-valued project can be jointly exploited by the players, but that a high-valued project and two medium-valued projects cannot, then $\tilde{\pi}(m)$ would depend on the parity of m and monotonicity would break.

6 Applied Insights

This section connects our theoretical analysis to two key literatures: buyer-supplier relationships and the persistent productivity differences across firms.

6.1 Buyer-Supplier Collaborations

The economics literature on buyer-supplier relationships has predominantly examined issues such as vertical integration in the presence of relationship-specific investments (Williamson, 1975; Grossman and Hart, 1986; Hart and Moore, 1990), optimal contracts under externalities or agency issues (see references in Tirole, 1988, Chapter 4), and, more recently, relational contracts for supplier allocation (Board, 2011; Andrews and Barron, 2016). While these studies justifiably assume predetermined gains from trade to address their specific objectives, our research explores a complementary direction: scenarios requiring collaborative experimentation to determine the gains from trade, often across multiple products or markets.

Our model formalizes the process of collaborative experimentation in buyer-supplier relationships through several key elements. The parameter m represents the number of product categories or market geographies, reflecting the multi-domain nature of buyer-supplier interactions. Both firms make non-contractible investments of c for experimentation. These investments are observable to both parties but not verifiable by third parties, hence not contractible. The innovation process involves both firms, each possessing complementary and indispensable expertise or resources. Even after the exploration phase, when parties agree on an input or service to exploit, non-contractible investments (also c) remain essential. These include efforts such as worker training and marketing. The distribution of benefits is asymmetric because final product proceeds accrue to the buyer (high α), who compensates the supplier through either the upfront transfer w or the bonus b.

Our theoretical analysis both draws from and contributes to an extensive body of case-study literature on buyer-supplier dynamics. This literature emphasizes experimentation and trust as critical factors for successful collaborations, particularly in contexts where benefits are asymmetrically distributed. 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). This asymmetry underscores the central role of trust, given the inherent limitations 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 primarily focuses on credibility trust, operating under the assumption that both parties desire collaboration. Consequently, in this section 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) offer insights into the dynamic nature of buyer-supplier relationships, emphasizing the central role of relational contracts in these interactions. They describe an initial "search and trial phase" that evolves into an "expansion phase," marked 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." The pervasiveness of gradualism is in line with our findings, particularly Proposition 4, which shows the potential optimality of gradually expanding collaborative scope, and supports our extension in Section 5.1 examining the strategic delay of high-risk ventures in these relationships.

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 observation supports our theoretical model, where increases in scope are driven by discrete "events" that change the players' continuation value from the collaboration, rather than the mere passage of time.

Lastly, our analysis, particularly Proposition 5, showed that the long-term scope of a collaboration is largely determined during the initial exploration phases, with early outcomes significantly influencing the trajectory and ultimate extent of the partnership. This finding is corroborated by the existing literature. 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.

6.2 Persistent Performance Differences

While much of our focus has been on interactions between firms, our model serves as a valuable lens for examining employer-employee dynamics. One can conceptualize one party in our model as the employer and the other as the employee, where, for instance, benefits consistently accrue to the employer. Furthermore, the different domains of collaboration can be seen as various dimensions of the production improvement process.

With this interpretation in mind, our work also contributes to the literature on persistent performance differences among seemingly similar enterprises (see Syverson, 2011; Gibbons and Henderson, 2013, and references therein). Numerous empirical studies have documented enduring disparities in firm performance across a range of industries, with these gaps proving surprisingly robust against plausible explanations such as market competition or local geographical and demand conditions, while being strongly associated with managerial practices (c.f. Bloom and Van Reenen, 2007). According to Gibbons and Henderson (2013), and the body of evidence they review, variations in managerial practices, because of their reliance on relational contracts, are key in creating productivity disparities across firms. We adapt for our purposes their categorization of explanations: (i) managers might either be unaware of their poor performance, or, even if aware, believe that the best practices from other firms are not suitable for their context; (ii) managers are aware of their poor performance and are able to seek superior managerial practices suitable to their context, but opt not to; and (iii) managers are "striving mightily" to adopt superior practices but face hurdles during the implementation phase.

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).[14] 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. Our analysis in Section 4.3 provides an alternative rationalization of explanation (ii) by showing that collaborations may not reach their maximum potential scope despite identical initial conditions. When starting with limited scope, players might prefer to cease exploration in early cooperation domains and switch to exploitation, rather than continue exploring for an extended period of time in hopes of identifying superior practices valuable enough to enable scope expansion.

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 strategies, operating under a discount factor where a gradual approach is optimal and the maximum potential scope of collaboration is guaranteed to be achieved. Their paths may diverge if one organization discovers a highly valuable practice in the initial domain early on, thus expanding its scope, while the other does not. The second organization, still attempting to achieve success in the first domain, 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.

[14] Our framework could be easily modified to rationalize the first explanation by introducing correlation between project benefits within each domain. If by chance the first projects explored in a domain are disappointing, the players increasingly believe no projects in the domain are profitable and may stop exploring and terminate their relationship. In contrast, if the early projects are valuable, the players may switch to exploitation and enjoy high long-run profits. Unsurprisingly, these dynamics would also arise in a single-agent context.

7 Concluding Remarks

This paper presents a framework for analyzing the dynamics of multi-domain collaborative experimentation in scenarios where benefits are unevenly distributed among participants and any experimentation policy must be self-enforcing. Our model yields three key insights. First, when the initial relationship value is low, the collaborating parties do not treat each domain of experimentation independently and they engage in extended exploration phases. Second, experimentation often progresses gradually, with parties initially exploring some domains and potentially expanding to others based on initial success, and exploration of all domains is not guaranteed. Third, cross-domain relational interdependence in optimal experimentation leads to seemingly counterintuitive exploration/exploitation decisions, including prolonged exploitation of ultimately discontinued projects or revival of previously abandoned ones.

Our analysis built upon a deliberately simple experimentation framework. This choice ensured that the first-best experimentation policy remained straightforward, making departures due to strategic considerations more striking. Future work could extend the current framework in several directions. For example, relaxing the assumption of identically and independently distributed project benefits within domains could help address questions related to directed innovation strategies and differentiate between radical and incremental innovation (c.f. Callander, 2011; Garfagnini and Strulovici, 2016; Callander and Matouschek, 2019). Further, we assumed that both players' cooperation was necessary for exploration and exploitation, keeping their outside options independent of experimentation. Future research could explore scenarios where players' outside options evolve based on their experimentation history, examining how this additional interdependence affects joint experimentation dynamics. Finally, introducing asymmetric roles in the collaboration presents another natural extension. One could model a scenario where exploration requires only one player (e.g., an R&D unit), while exploitation needs a different player (e.g., a Sales unit). This approach would enable analysis of cooperation dynamics in contexts where exploration and exploitation efforts are disentangled (see Krieger et al., 2019; Lizzeri et al., 2024, for qualitative and theoretical treatments, respectively).

References

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

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

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.

Bonatti, Alessandro and Johannes Hörner, "Collaborating," *American Economic Review*, 2011, *101* (2), 632–663.

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 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.

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.

Ghosh, Parikshit and Debraj Ray, "Cooperation in Community Interaction without Information Flows," *The Review of Economic Studies*, 1996, *63* (3), 491–519.

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

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. Forthcoming.

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

Kranton, Rachel E., "The Formation of Cooperative Relationships," *Journal of Law, Economics, & Organization*, 1996, *12* (1), 214–233.

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.

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.

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

Reshidi, Pëllumb, Alessandro Lizzeri, Leeat Yariv, Jimmy Chan, and Wing Suen, "Sequential Sampling by Individuals and Groups: An Experimental Study," Technical Report, Technical Report, Working paper 2024.

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

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.

— , *A Theory of Corporate Finance*, Princeton University Press, 2006.

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.

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.

Appendix

Proof of Proposition 1.

Since there are no interdependencies across domains, player 0 treats each domain independently and identically. Further, given that there exists an infinite number of ex-ante identical projects in each domain, the optimal experimentation policy conditions only on the project with the highest value amongst all previously explored projects, whose value we denote $\hat{v}$. One can write the Bellman Equation for player 0:

$$B^0(\hat{v}) = \max_{\text{explore, exploit } \hat{v}} \left\{ \mathbb{E}\left(s'\right) - 2c + \delta \mathbb{E}\left(B^0\left(\max\left(\hat{v}, s'\right)\right)\right), \hat{v} - 2c + \delta B^0\left(\hat{v}\right) \right\}. \quad (8)$$

The first term in the maximum operator corresponds to the player's expected surplus when exploring one more project (chosen at random, since all unexplored projects are ex ante identical) 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, for any $\delta < 1$, one can use Blackwell's Sufficient Conditions to show that there exists a unique solution to the Bellman Equation, and hence the threshold rule dictated by v^0 is a solution. Finally, this threshold is determined by:

$$\frac{1}{1-\delta}(v^0 - 2c) = \mathbb{E}(v_p - 2c) + \frac{\delta}{1-\delta}\mathbb{E}(\max\{v, v^0\} - 2c), \tag{9}$$

where standard comparative statics arguments imply that v^0 is increasing in δ. □

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:

$$c\sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j \geq v^0} + \max\{0, c - (1-\alpha)\mathbb{E}\left(v_p\right)\}\sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j < v^0} \leq \delta \sum_{j=1}^{m} \mathcal{C}(\hat{v}'_j), \tag{10}$$

$\forall(\hat{v}'_1, \ldots, \hat{v}'_m) \geq (\hat{v}_1, \ldots, \hat{v}_m)$, which corresponds to (2) when the players implement the first-best policy and where $\mathcal{C}(\hat{v}'_j)$ denotes the continuation value in domain j under the first-best policy. Note that $\mathcal{C}(\hat{v}'_j)$ (i) is constant below v^0, (ii) is such that $\lim_{x\uparrow v^0} \mathcal{C}(x) > \lim_{x\downarrow v^0} \mathcal{C}(x)$ and (iii) is increasing above v^0. Given such properties, setting $\hat{v}'_i = \max\{\hat{v}_i, v^0\}$ both minimizes the right-hand side and maximizes the left-hand side of (10). Thus, an equivalent condition is:

$$m \cdot c \leq \delta\Big(\sum_{j=1}^{m} \frac{1}{1-\delta}(\max\{\hat{v}_j, v^0\} - 2c)\Big), \tag{11}$$

which corresponds to the expression stated in the proposition. Finally, the existence of a threshold δ^0 was proven in the text. □

Proof of Corollary 2. This result was proven in the text. □

Proof of Proposition 4. We first prove the existence of δ^*. Suppose $\delta_1 < \delta_2$ and, by contradiction, that the 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 maximal if and only if the continuation value at date 1 exceeds $2 \cdot \max\{0, c-(1-\alpha)\mathbb{E}(v_p)\}$. 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, by the same argument as above, when $\delta \to 1$, $\mathcal{C}(\cdot) \to \infty$, implying maximal scope. As a result, $\bar{\delta} < 1$.

As discussed in the text, $\delta^* \leq \bar{\delta}$ because any maximal relational contract is non-empty. We next show that this inequality need not be strict. In doing so, we set $\alpha = 1$, and by continuity these arguments will extend to an open set of parameters.

To show that $\delta^* = \bar{\delta}$, we consider a three-point support of benefits: $\{0, \underline{v}, \bar{v}\}$. To simplify the continuation values, we will assume $\mathbb{E}(v_p) = 2c$ and list the necessary inequalities for $\delta^* = \bar{\delta}$ below. However, as all these inequalities hold strictly, upon increasing the expected project value slightly or considering a continuous-distribution approximation, all the inequalities presented will continue to hold. Throughout, $\Pr(\cdot)$ denotes the probability of a given realization. These inequalities are:

$$\frac{\underline{v} - 2c}{1-\delta} > \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} \tag{12}$$

$$2c < \frac{\delta}{1-\delta}(\bar{v} - 2c) \tag{13}$$

$$c > \frac{\delta}{1-\delta}(\underline{v} - 2c) \tag{14}$$

$$2c > \frac{\delta}{1-\delta}(\underline{v} - 2c) + \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} \tag{15}$$

$$c > \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} + \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}\mathrm{npv}^0 \tag{16}$$

$$\begin{aligned} 2c < {} & \frac{2\Pr(\bar{v})(1-\Pr(\bar{v}))}{1-\delta\big(1-2\Pr(\bar{v})(1-\Pr(\bar{v}))\big)}\Big((\bar{v}-2c)\frac{\delta}{1-\delta} + \mathrm{npv}^0\Big) \\ & + \frac{\Pr(\bar{v})^2}{1-\delta\big(1-\Pr(\bar{v})^2\big)}\frac{\delta}{1-\delta}2(\bar{v}-2c) \end{aligned} \tag{17}$$

$$\mathrm{npv}^0 = \frac{\delta}{1-\delta}\frac{\Big(\Pr(\bar{v}) + \Pr(\underline{v})\Big)\Big(\frac{\Pr(\bar{v})\bar{v}}{\Pr(\bar{v})+\Pr(\underline{v})} + \frac{\Pr(\underline{v})\underline{v}}{\Pr(\bar{v})+\Pr(\underline{v})} - 2c\Big)}{1-\delta(1-\Pr(\bar{v})-\Pr(\underline{v}))} \tag{18}$$

Inequality (12) ensures that player 0 prefers to exploit $\underline{v}$. Inequality (13) ensures that the players can implement the first-best policy upon discovering $\bar{v}$. Inequality (14) ensures that the players are unable to exploit a project worth $\underline{v}$ in isolation and, by extension, cannot jointly exploit two such projects. Inequality (15) implies that the players cannot exploit a project worth $\underline{v}$ while exploring along the additional domain. Inequality (16) implies that the players would be unable to begin exploring if the players began their exploration on one domain and, upon finding a project worth $\bar{v}$, started exploring the additional domain.[15] The continuation value under this experimentation policy is bounded above by npv^0: the net-present value of a single domain under the first-best experimentation policy, the value of which is stated in (18). Finally, Inequality (17) is a necessary condition for the maximal policy to be feasible at date 1. This condition is necessary, but not sufficient, as the continuation value is computed assuming that if the players discover one project worth $\bar{v}$ before doing so on the other domain, the highest-valued project on the other domain is zero. One can then use Mathematica to show that these constraints jointly hold strictly.[16]

All that remains to show is that $\delta^* < \bar{\delta}$ may also occur. To prove this inequality, consider a binary support for the distribution of valuations any by a similar argument, if the conditions hold strictly with this distribution of valuations, the conditions will hold with a continuous approximation. With a binary distribution, Equations (4) and (5) are identical. As a result, gradual is always feasible for a strictly larger set of discount factors.

Finally, if, by contradiction, $|\mathbf{P}^1| \neq 0$ and $|\mathbf{P}^t|$ never equals 2, $|\mathbf{P}^t| = 1 \forall t$. In this case, the players might as well only collaborate in one of the two domains. However, the players can replicate this single-domain experimentation policy on the other domain, thereby doubling their expected payoff. $\square$

The following lemma is useful in proving the remaining propositions.

Lemma 1 (**Properties of the Second-Best Experimentation Policy**)
Consider $m = 2$ and suppose that non-empty optimal relational contracts exist. Let t be a period in which $\hat{v}_1 \geq \hat{v}_2$ and Inequality (3) does not hold. Then, in any optimal relational contract, $|\mathbf{P}^t| \geq 1$. Furthermore, without loss of optimality, the players' actions in period t satisfy one of the following:

[15] Further, it will never be optimal to explore the additional domain upon finding $\underline{v}$, since such a project cannot be exploited.

[16] Code available upon request.

1. *If* $|\mathbf{P}^t| = 2$*, the players either (i) explore both domains or (ii) explore domain 2 and exploit in domain 1.*

2. *If* $|\mathbf{P}^t| = 1$*, the players either (i) exploit in domain 1 or (ii) explore domain 2.*

Proof of Lemma 1. In each domain j, the players can either (i) explore a project, (ii) exploit the project with value $\hat{v}_j$, or (iii) neither explore nor exploit any project. Consequently, there are nine possible combinations of actions across the two domains.

We prove that $|\mathbf{P}^t| \geq 1$. Given a non-empty optimal relational contract, there exists an earliest date $t' \geq 1$ where $|\mathbf{P}^{t'}| \neq 0$. $t' > 1$ is impossible, as implementing the same experimentation policy $\hat{\mathbf{P}}(\cdot)$ starting at date $t = 1$, which would be strictly better due to discounting. If $|\mathbf{P}^1| \neq 0$, then $|\mathbf{P}^t| \neq 0$ for all subsequent t, as players can replicate date 1 payoffs and continuation value. Thus, $|\mathbf{P}^t| \geq 1$ in period t, eliminating one of the nine possible combinations of actions.

Next, since (3) does not hold under the first-best experimentation policy, it follows from Corollary 1 that the players can exploit at most one project in the current period. Furthermore, conditional on exploiting only one project, they will exploit the project with the highest value, $\hat{v}_1$. This observation eliminates an additional three of the nine possible combinations of actions.

Finally, if the players explore only one domain without exploiting a project in the other, they prioritize domain 2. While exploration yields identical current-period payoffs across domains, a higher-valued project in domain 2 generates a weakly higher continuation value since $\hat{v}_1 \geq \hat{v}_2$. This eliminates one more action combination, leaving the four listed in the lemma. □

Proof of Statement (ii) in Proposition 5. The proof of result (i) is proved in the text. We consider a three-point support for the benefits v_p of $0 < \underline{v} < \bar{v}$ and $\alpha = 1$. As all the inequalities regarding the optimality of the probabilistically bounded experimentation policy stated below hold strictly, these results will continue to hold when considering either a continuous approximation for this distribution of benefits and/or

$\alpha < 1$. We list all the inequalities and comment on each one separately below.

$$2c < \frac{\delta}{1-\delta}(\bar{v} - 2c) + \mathcal{C}^0(\text{explore}) \tag{19}$$

$$c < \frac{\delta}{1-\delta}(\underline{v} - 2c) \tag{20}$$

$$2c > \frac{\delta}{1-\delta}(\underline{v} - 2c) + \mathcal{C}^0(\text{explore}) \tag{21}$$

$$2c > \mathcal{C}^0(\text{explore}) + \frac{\delta}{1-\delta}\Big(\Pr(\bar{v})\bar{v} + (1-\Pr(\bar{v}))\underline{v} - 2c\Big) \tag{22}$$

$$\begin{aligned}\frac{\underline{v} - 2c}{1-\delta} > v &:= \mathbb{E}(v_p - 2c) + \frac{\delta}{1-\delta}\Big(\Pr(\bar{v})(\bar{v} - 2c) + \Pr(\underline{v})(\underline{v} - 2c)\Big) \\ &+ \Big(\Pr(\bar{v}) + \Pr(\underline{v})\Big)\frac{\delta}{1-\delta}(\underline{v} - 2c) + \Big(1 - \Pr(\bar{v}) + \Pr(\underline{v})\Big)\delta v\end{aligned} \tag{23}$$

$$c \leq \mathcal{C}^0(\text{explore}) + \frac{\delta \Pr(\bar{v})\mathcal{C}^0(\text{explore})}{1 - \delta(1 - \Pr(\bar{v}))} \tag{24}$$

Inequality (19) implies that the players can replicate the first-best experimentation policy upon discovering a project with value $\bar{v}$, where $\mathcal{C}^0(\text{explore})$ denotes the continuation value of exploration associated with player 0. Inequality (20) ensures that the players are able to exploit a project worth $\underline{v}$ in isolation and, by extension, can jointly exploit two such projects. Inequality (21) ensures that the players cannot expand their scope to two while exploiting the project worth v^0. This inequality uses $\mathcal{C}^0(\text{explore})$ 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 (which Inequality (20) ensures they can), explore a project on the other domain while maintaining a scope of 1, or conduct 2 explorations. Inequality (22) ensures that conducting two explorations is not feasible (i) because the upper-bound on the continuation value for the new domain is provided by the first-best benchmark and (ii) because the upper bound on the domain with a project with value $\underline{v}$ is also provided by the first-best benchmark after the current period. Next, Inequality (23) ensures that the players prefer to exploit the project worth $\underline{v}$ as opposed to (i) exploring the domain where the best project is worth 0 until discovering a project worth $\underline{v}$ or $\bar{v}$ and subsequently permanently exploiting this project and (ii) recalling the project on the other domain worth $\underline{v}$ and permanently exploiting it. These constraints imply that the players will remain bounded at $\underline{v}$ if the players reach this outcome. Finally, Inequality (24) ensures that the players' relational contract is non-empty if they conduct efficient exploration on the first domain, and conduct efficient exploration on the second domain if and only if the players discover a project worth $\bar{v}$ on the first domain. One can check that all constraints along with $\mathbb{E}(v_p) \geq 2c$ and the fact that player 0 prefers exploiting a project worth $\underline{v}$ to exploration jointly hold. For instance, upon setting $c = 1$ and $\Pr(\underline{v}) = .1$, one can use Mathematica to show that these inequalities hold strictly.[17] $\square$

Proof of Proposition 6. For the proof of both results consider $\alpha = 1/2$ and a value of δ where the first-best is unachievable.

Statement 1: As discussed in the text, it is sufficient to show that there exists an $\epsilon > 0$ such that if both projects' values fall in $(\tilde{v} - \epsilon, \tilde{v})$, the players exploit the better of the two projects and explore a project on the other domain.

By Lemma 1 and the fact that exploration is an equilibrium of the stage game when $\alpha = 1/2$, we know the players either exploit the higher-valued project in one domain and explore in the other or explore projects in both domains. Note that there exists a sufficiently small ϵ_1 for which the former experimentation policy is implementable. To see why, note that, by definition of $\tilde{v}$, the players can exploit both projects if both projects exceed $\tilde{v}$. If the project values fall slightly below $\tilde{v}$ and the players exploit the better of the two projects while exploring the second domain in all periods, the left-hand side of (2) is equal to c while the right-hand side is strictly greater than c (since the continuation value from exploration is strictly positive and that of exploiting a project with value arbitrarily close to $\tilde{v}$ is arbitrarily close to c). Thus, the question turns to whether the players prefer two explorations to one exploration and exploitation.

If the probability of finding a project with value exceeding $\tilde{v} - \epsilon$ is arbitrarily small, the net-present value from exploring projects on both domains approaches $2\mathbb{E}(v_p - 2c)/(1 - \delta)$. However, the net-present value from exploiting the better project and exploring another approaches $\mathbb{E}(v_p - 2c)/(1 - \delta) + (\tilde{v} - \epsilon - 2c)/(1 - \delta)$. As $\tilde{v} - \epsilon > v^0 > \mathbb{E}(v_p)$, the players would find it in their interest to exploit the better project, completing the argument.[18]

Statement 2: As exploration is an equilibrium of the stage game, the players conduct two explorations in period 1 and with positive probability identify two

[17] The code can be provided upon request.

[18] As seen in Figure 1b, these dynamics can be shown to happen without the limit arguments provided in this proof.

projects with values belonging to an arbitrarily small range around $v^0 + \epsilon$ and $v^0 - \epsilon$. There exists a δ sufficiently small such that neither project can be exploited. The players are unable to exploit either project in period 2 and, thus, must explore two new projects. Because the distribution of v_p is unbounded, for any δ, there exists a realization of v_p large enough such that $h(v^0 + \epsilon, v_p) > \tilde{v}$. Finally, in this region, the players follow the first-best experimentation policy and thus permanently exploit both projects. Therefore, with positive probability, the players exploit a project they have previously chosen not to exploit.

□