Multi-Project Collaborations

Charles Angelucci and Roi Orzach*

February 17, 2025

Abstract

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

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

JEL Classification Numbers: D21, D70, D83, L25

*Angelucci: MIT Sloan School of Management, 50 Memorial Drive, Cambridge MA 02142 (e-mail: cangeluc@mit.edu); Orzach: Department of Economics, MIT, 50 Memorial Drive, Cambridge MA 02142 (e-mail: orzach@mit.edu).

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

1 Introduction

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

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

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

We focus on Subgame Perfect Equilibria (relational contracts) that maximize the players' discounted cumulative joint payoffs (their "surplus"). As a first benchmark, Proposition 1 examines the single-player scenario, providing a straightforward solution in which 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. We refer to this optimal policy for the single-player scenario as the "first-best experimentation policy." Notably, this first-best policy would be optimal for the two players if all projects benefited them equally.

However, because experimentation in our setting requires both players' participation, asymmetric project benefits create incentive challenges. Our analysis centers on this scenario by assuming each project benefits only one player. The beneficiary's identity is revealed when players first cooperate on a project and is independently and identically distributed across projects. These asymmetric benefits create the key friction that may impede first-best experimentation, as implementing this policy requires credible promises of transfers between players. Such promises may lack credibility when players making transfers have insufficient continuation value in the collaboration. As mentioned above, since a player's continuation value equals the sum of continuation values across all domains of cooperation, experimentation choices in one domain affect all others.

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

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

Next, we analyze the second-best experimentation policy, which arises when the first-best policy is not implementable in the current period. We first examine the players' exploration and exploitation decisions, abstracting from the number of domains they engage in. Due to cross-domain interdependencies, the player's exploitation criterion becomes dynamic. Nonetheless, Proposition 4 shows that, unlike the first-best policy where explored projects are either permanently exploited or never used, the second-best policy is such that, with strictly positive probability, players (i) exploit projects temporarily or (ii) exploit previously unexplored projects rather than the most recently explored ones.

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

In Section 5, we examine how the potential scope of experimentation impacts its feasibility and profitability, drawing connections to the seminal work of Bernheim and Whinston (1990) on multilateral interactions. Further, we discuss extensions of the model included in the Online Appendix, in which the domains of cooperation are asymmetric or exhibit technological interdependencies.

Section 6 connects our main findings to the existing applied literature on buyer-supplier relationships 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), contributing to the strand that examines strategic interactions.[1] Bolton and Harris (1999) and Keller et al. (2005) consider settings in which players free-ride on each others' experimentation (see Hörner et al., 2022, for more recent work on this topic). Further, Liu and Wong (2023) consider an environment in which players compete to explore alternatives. In Strulovici (2010), players vote between a safe and a risky arm, with its asymmetric benefits revealed over time through experimentation (see also Anesi and Bowen, 2021). Further, Albrecht et al. (2010) examine a search problem where a committee determines which project to exploit. Chan et al. (2018) and Reshidi et al. (2025) compare group and individual decision-making, examining the effects of static vs. sequential information acquisition and voting rules. In contrast to these papers, our setting allows for voluntary transfers and requires the players to cooperate for both the exploration and exploitation of projects. Most significantly, players experiment simultaneously across multiple domains.

Multi-domain experimentation poses analytical challenges. As noted in Bergemann and Välimäki (2008),"it is well known that [a Gittins] index characterization is not possible when the decision maker must or can select more than a single arm at each t," due to the optimality of recalling past projects.[2] With infinitely many ex ante identical projects, as we assume, project recall is absent and a Gittins index exists in the single-agent case (Bergemann and Välimäki, 2001). In our setting, cooperation between players is needed to experiment and project benefits are asymmetric. Therefore, an experimentation policy must form an equilibrium and we demonstrate that this requirement generates rich dynamics—such as project recall, temporary exploitation, and idling on previously explored domains—even though these very dynamics complicate the problem by precluding a Gittins representation.

Secondly, this work relates to the literature on relational contracts (see e.g., Bull, 1987; Macleod and Malcomson, 1989; Baker et al., 1994, 2002; Levin, 2002, 2003, for early contributions).[3] Halac (2014) studies a setting in which the value of the players'

[1]Our setting resembles standard search problems by modeling many alternatives for players to explore. However, unlike typical search problems where rewards come only at the end from the best explored alternative, our model allows players to benefit each time they cooperate on a project, without settling on one. For this reason, we use the broader term "experimentation" rather than "search." Moreover, existing models of strategic experimentation with bandits often limit options to a few alternatives, like a risky and a safe project. We assume an infinite number of i.i.d. projects to eliminate aggregate uncertainty, making the dynamics driven purely by strategic factors.

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

[3]Also at the intersection of the bandit and the relational contracting literatures, Urgun (2021)

relationship increases exogenously with its duration, allowing for greater efficiency. In our setting, players' experimentation endogenously shapes the continuation value of their relationship, which may not increase monotonically over time. For instance, while exploration in any domain enhances the players' continuation value, exploitation diminishes it, potentially hindering experimentation in other domains.

A closely related paper to ours is Chassang (2010). In his model, the agent knows which arms are productive and which are not, while the principal, at the outset, 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–apt for modeling firms–removes the need for inefficient on-path punishments. Yet, it generates dynamics similar to those observed in collaborations between and within firms (see Section 6.2).[4]

Finally, we contribute to the literature 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. They 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 players' experimentation can expand or contract over time due to the evolving continuation value of the relationship. Moreover, the two settings make opposite predictions about how the discount factor affects players' incentives to "start small.' In our setting, a higher discount factor reduces this need, whereas in the frameworks analyzed by Watson (1999, 2002) and the broader dynamic screening literature (e.g., Ely and Välimäki, 2003; Acharya and Ortner, 2022), a higher discount factor increases it, as separation becomes harder to achieve.

---

examines a scenario where a principal interacts with multiple agents whose publicly-observable types depend on the contracting history.

[4]Introducing money in Chassang (2010), where information asymmetry plays a central role, would make the value of the relationship constant on path. We further discuss Chassang (2010) in Footnote 13 and Section 6.2. For a setting similar to Chassang (2010) but with imperfect transfers and uncertainty about the value of the relationship, see Venables (2013). For work on experimentation in principal-agent settings with commitment, see Halac et al. (2016) and Ide (2024).

2 The Setup

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

Each project in $\mathbf{P}^t$ costs $c > 0$ for each player and has initially unknown time-invariant value $v_p \in \mathcal{R}$, which is publicly observed after the first cooperation. We assume that for each project, a single player receives the entire value v_p of the project.[5] 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 $1 - \alpha$.

We say that a project is being "explored" when cooperated on for the first time and "exploited" when cooperated on in both the current period and at least one prior period. There are no intertemporal restrictions on project availability.

We make two assumptions on the distribution of project values. First, we assume that the distribution of v_p admits a continuous density with a convex support equal to $\mathcal{R}^+$. Next, we assume $\mathbb{E}(v_p) \geq 2c$. These assumptions ensure that the first-best experimentation policy will be unique and non-empty.[6]

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 $w_{i,-i}^t \in \mathcal{R}^+$ denotes such a transfer from player i to player $-i$. At the end of each period t, players

[5]Results hold as long as, for each project, one player values it above c and the other below it.

[6]Assuming an unbounded support also simplifies some technical aspects of the proofs. Further, $\mathbb{E}(v_p) < 2c$ 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 factor is key.

again make discretionary transfers to each other, where $b^t_{i,-i} \in \mathcal{R}^+$ denotes such a transfer from player i to player $-i$.[7] Finally, player i's period t payoff is equal to:

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

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

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

Relational Contracts. A relational contract is a complete plan for the relationship. Let $h^t = (\mathbf{w}^1, \mathbf{P}^1, \mathbf{v}^1, \mathbf{x}^1, \mathbf{b}^1, \ldots, \ldots, \mathbf{w}^{t-1}, \mathbf{P}^{t-1}, \mathbf{v}^{t-1}, \mathbf{x}^{t-1}, \mathbf{b}^{t-1})$ denote the history up to date t and $\mathcal{H}^t$ the set of possible date t histories, where boldface lowercase letters indicate vectors. Then, for each date t and every history $h^t \in \mathcal{H}^t$, a relational contract describes: (i) the $\mathbf{w}^t$ transfers; (ii) the set of projects $\mathbf{P}^t(\mathbf{w}^t)$ as a function of $\mathbf{w}^t$; and (iii) the $\mathbf{b}^t(\mathbf{w}^t, \mathbf{P}^t, \mathbf{v}^t, \mathbf{x}^t)$ transfers as a function of $\mathbf{w}^t$, $\mathbf{P}^t$, and the realizations of $\mathbf{v}^t$ and $\mathbf{x}^t$. A relational contract is self-enforcing if it constitutes a Subgame Perfect Equilibrium of the repeated game. Within this class, we focus on equilibria that maximize joint surplus. Restricting to pure strategy equilibria is without loss of optimality since (i) mixing on transfers only increases the maximal transfers players can promise 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 $P^t_i = \emptyset$ and $b^t_{i,-i} = 0$ if these choices have not been made yet and (ii) by permanently breaking off their relationship (i.e., reverting to the worst equilibrium of the stage game from the next period onward). This punishment is without loss of optimality as it occurs out-of-equilibrium (c.f. Abreu, 1986).[8] Throughout, a relational contract is defined as "non-empty" if $\Pr(\sum_t |\mathbf{P}^t| > 0) > 0$.

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

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

3 First-Best Experimentation

We characterize the optimal experimentation policy for a benchmark where a single decision maker, "player 0," maximizes the sum of the payoffs of both players. This optimal experimentation policy is identical to the one we would obtain if we modified the model described in Section 2 so that the projects always benefit both players equally. The proof of the following proposition closely follows Bergemann and Välimäki (2001) and is provided in the Appendix, along with proofs for all other statements omitted from the main text.

Proposition 1 (**First-Best Experimentation Policy**)
For each domain j and period t, player 0 adopts the following experimentation policy: if a previously-explored project p has the highest value and $v_p \geq v^0(\delta)$, exploit it; 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 he chose not to exploit in the past. Finally, as the discount factor increases, the value of exploration increases, which explains the comparative statics result for v^0.

In summary, the first-best policy maximizes experimentation scope, with exploration/exploitation in each domain following a fixed, time-invariant threshold. We now analyze the model from Section 2, identifying when these features break down and the resulting dynamics.

4 Main Analysis

In Section 4.1, we characterize the class of optimal relational contracts on which the analysis focuses and establish a necessary and sufficient condition for an experimentation policy to be implementable by an optimal relational contract. In Section

4.2, we provide the conditions under which the players can implement the first-best policy. In Section 4.3, we characterize key properties of the optimal experimentation policy when they are unable to implement the first-best policy.

4.1 Optimal Experimentation Policies: Implementability

In our setting, surplus-maximizing relational contracts depend on the players' beliefs about the projects, which we denote 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 optimal. The following proposition formalizes this characterization and provides a necessary and sufficient condition for an experimentation policy $\hat{\mathbf{P}} : \{\Delta(v_p, x_p)\}_{p\in\mathcal{P}} \to \mathcal{P}$ to be implementable by an optimal relational contract.

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

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

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

The optimal experimentation policy in any given period depends only on the values of the most valuable projects identified in each of the m domains, denoted by $\hat{v}_1, \ldots, \hat{v}_m$, where $\hat{v}_j := 0$ if no projects have been explored in domain j. Players never exploit a project with a lower value than another, as doing so would reduce their joint payoff and make Inequality (2) (weakly) tighter. Thus, tracking $\hat{\mathbf{v}} := (\hat{v}_1, \ldots, \hat{v}_m)$ is sufficient to represent players' beliefs about the projects. For each j, they choose one of three actions: remain idle ($a_j = 0$), explore a new project ($a_j = 1$), or exploit the highest-valued known project ($a_j = 2$). The experimentation policy is then determined by solving the following Bellman equation, where $B(\hat{\mathbf{v}})$ represents the players' joint surplus:

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

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

Notably, Inequality (4) aggregates incentives across domains, introducing interdependencies. The implications of these interdependencies for players' experimentation will be the focus of our analysis. Moreover, they prevent an analytical characterization of the optimal policy, as we explain below.

Further, we caution against the following intuition. While the players' joint surplus, $B(\cdot)$, increases over time and the continuation value for a fixed policy, $\mathcal{C}(\mathbf{a}, \cdot)$, also grows, the equilibrium continuation value, $\mathcal{C}(\mathbf{a}, \cdot)$, is *not* necessarily monotonic. This non-monotonicity arises even under the first-best policy described in Proposition 1. For instance, $\mathcal{C}(a(v^0+\epsilon), v^0+\epsilon) < \mathcal{C}(a(0), 0)$, because after identifying a project with a value slightly above v^0, player 0 becomes nearly indifferent between exploiting the current project and continuing to explore. This implies that the continuation value associated with exploration exceeds that of exploitation. The non-monotonic nature of the continuation value further complicates the analysis, as Inequality (4) does not necessarily relax over time.

4.1.1 Challenges in Characterizing Optimal Experimentation

As discussed in Section 1.1, our setting does not admit a Gittins Index characterization. More generally, any characterization of the optimal experimentation policy is generally infeasible. First, the choice set is discrete, which precludes the use of continuous optimization methods. Second, due to Inequality (4), this multi-dimensional optimization problem cannot be decomposed into m independent optimization problems. As a result, the curse of dimensionality arises for $m > 1$ due to two interrelated reasons. First, even for $m = 2$, the choice set in any given period t consists of 9 options (or 5, under symmetry), and this number grows exponentially with m. Second, determining whether a given choice is feasible and optimal requires knowledge of $B(\hat{\mathbf{v}}')$ for all $\hat{\mathbf{v}}' \geq \hat{\mathbf{v}}$, and subsequently, computing its respective integral over all possible future values of $\hat{\mathbf{v}}'$ for each choice $\mathbf{a}$ to evaluate $\mathbf{C}(\mathbf{a}, \hat{\mathbf{v}})$. If the support of

v_p were discrete with cardinality n, the problem could, in principle, be solved using "backward induction" on the Bellman equation. However, this approach is analytically feasible only when both n and m are very small (in Online Appendix A we provide a characterization for the $n = m = 2$ case).[9]

4.2 Implementability of First-Best Experimentation

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

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

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

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

When Inequality (5) is satisfied, the continuation value of the relationship is sufficiently high to enable the implementation of the first-best experimentation policy.

[9]Note that meaningful exploration/exploitation decisions require the support of v_p to have a cardinality strictly greater than 2. For this reason, we assume a continuous support, which also facilitates the presentation of some of our results. However, it follows from Lemma 1 in the Appendix that none of our results rely on continuous supports.

Because the players can pool relational incentives across domains, 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 $h(\hat{v}_1, \dots, \hat{v}_m)$ is not the arithmetic mean of the values $\hat{v}_1, \dots, \hat{v}_m$ 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 (5) 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.[10]

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

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

Proof of Corollary 1. Proposition 3 establishes that these conditions are jointly sufficient. Fixing $\hat{\mathbf{v}}$, the continuation value associated with permanent exploitation of $\hat{\mathbf{v}}$ is weakly lower than the continuation value under the first-best policy at $\hat{\mathbf{v}}$. Hence, if the players are able to permanently exploit $\hat{\mathbf{v}}$, they can also implement the first-best experimentation policy. This implies that these conditions are not only sufficient but also jointly necessary. □

The conditions stated in Corollary 1 imply that, in expectation, the players achieve the permanent exploitation outcome weakly later than if they could follow the first-best experimentation policy from period 1 onward. This delay relative to the first-best is strictly positive when $\delta < \delta^0$. In fact, as we will show in Proposition 6, permanent exploitation in all domains of cooperation is not 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 (and the optimal relational contract is non-empty), the players face a simple decision in each

[10] 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.

period: either to exploit the best project found thus far or to explore a new project. The exploitation threshold in this setting is time-invariant, as the players' continuation value depends solely on the value of the best project in this single domain.

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

In this subsection, we have provided the conditions on the best projects found in each domain under which the players implement the first-best experimentation policy. 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. We refer to experimentation in this region as "second-best experimentation." A non-empty region where the second-best policy is relevant (i.e., $\delta \geq \delta^*$) but the first-best policy is not implementable (i.e., $\delta < \delta^0$) follows from Corollary 2 and is further examined in Section 5.1. This analysis focuses on the case where the maximal potential scope of experimentation, m, is strictly greater than 1 (for the case $m = 1$, see Corollary 2).

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

4.3.1 The Dynamics of Exploration-Exploitation Decisions

Under the first-best policy, each domain is treated independently and identically, with a time-invariant threshold for project exploitation. This time-invariance ensures that once a project is exploited or deemed unworthy of exploitation, the decision is permanent. For collaborative experimentation, players aggregate incentives across all domains, with domains being treated neither identically nor independently. We show that this observation implies that the criterion used to determine project exploitation is dynamic. As a result, the players may exploit a project temporarily, and further, they may recall a project they previously chose not to exploit.

Proposition 4 (**Temporary Exploitation and Recall of Projects**)
When the players cannot implement the first-best experimentation policy in period 1 *and the optimal experimentation policy is non-empty (i.e., when $\delta \in [\delta^*, \delta^0)$), then with strictly positive probability for any $m > 1$, at least one of the following occurs:*

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

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

The first statement 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' scope of experimentation 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. In Section 5.2, we simulate the optimal experimentation policy for a parameterized example to further illustrate and develop intuition about the emergence of this behavior.

To understand the intuition behind the second statement, consider a scenario where the discount factor δ is small enough to prevent the exploitation of projects with values only slightly above the threshold v^0. Suppose the players' scope of experimentation is maximal, which occurs, for instance, when $\alpha = 1/2$.[11] If period 1 explorations yield two projects with values just above v^0, the players must explore again in the next period. However, if a newly explored project has a sufficiently high value, it can raise the continuation value of their relationship, potentially enabling first-best experimentation. In this case, they may optimally revert to exploiting a period 1 project despite initially choosing to explore further.

Temporary project exploitation or project recall are common in experimentation settings, and can arise due to various factors, including the presence of a finite number of projects or project characteristics that may not be fully revealed immediately. Our analysis shows that strategic interactions alone can also drive these behaviors.

4.3.2 The Dynamics of the Scope of Experimentation

Proposition 3 established a threshold δ^0, such that when $\delta \geq \delta^0$, players implement the first-best policy starting in period 1, maintaining maximal scope. We now examine the dynamics of the players' scope of experimentation when $\delta \in [\delta^*, \delta^0)$ and show that scope is not always maximal along the equilibrium path. To focus on the relevant case, we assume $(1-\alpha)\mathbb{E}(v_p) < c$, requiring player 1 to incentivize player 2 to explore. If instead $(1-\alpha)\mathbb{E}(v_p) \geq c$, project exploration is a static equilibrium, and optimal experimentation always maintains maximal scope.

We define a non-empty experimentation policy as "initially maximal" if $|\mathbf{P}^1| = m$, "initially limited" if $|\mathbf{P}^1| < m$, "terminally maximal" if $\lim |\mathbf{P}^t| = m$, and "terminally limited" if $\lim |\mathbf{P}^t| < m$. An initially maximal policy is always preferred over an initially limited one whenever both are implementable, as exploring all domains provides immediate benefits ($\mathbb{E}(v_p) \geq 2c$) and maximizes the continuation value of the relationship. The key question, then, is whether an initially limited policy can be implemented when an initially maximal one cannot. Intuitively, starting with a limited number of domains and allowing for future expansion may be more sustainable, as (i) it reduces early reneging temptation while maintaining a high continuation value due to these potential future scope expansions, and (ii) finding valuable projects in early domains can enable both their exploitation and the exploration of additional domains. We show that this intuition holds when the maximum potential scope of

[11] When $\alpha = 1/2$, exploration occurs in each domain in the static equilibrium, so any optimal relational contract implements an experimentation policy with maximal scope throughout.

experimentation m exceeds a threshold, but may fail below it.

To build intuition, we present the period-1 version of Inequality (2) for a specific initially limited policy, where players explore projects in domain 1 during period 1, and the corresponding inequality for the initially maximal policy, respectively:

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

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

where $B(\cdot)$ was defined in Equation (3), and where we explicitly highlight the relationship between $B(\cdot)$ and δ, as this will play a key role in the intuition below. We focus solely on period 1, remaining agnostic about the long-term dynamics of both policies. We note that the right-hand side of (7) increases with δ, indicating the existence of a cutoff $\bar{\delta}(m) \in (0, 1)$ below which this constraint is violated. Therefore, the question is whether (6) holds for $\delta < \bar{\delta}(m)$.

We proceed under the (incorrect) assumption that $B(\cdot|\delta)$ is continuous with respect to δ.[12] Under this assumption, and using (6) and (7), the initially limited policy outlined above is optimal when δ is just below $\bar{\delta}$ if and only if:

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

The right-hand side of this inequality represents the average surplus per domain from period 2 onward under the initially maximal policy, and, intuitively, is bounded above by that of a single domain under the first-best policy. Conversely, the left-hand side represents the total surplus across domains from period 2 onward under the initially limited policy. By monotonicity of the Bellman equation, the left-hand side of Inequality (8) is bounded below by:

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

where the last step follows from Inequality (2). Since this lower bound diverges with m, Inequality (8) holds for sufficiently large m, implying that an initially limited policy has a lower critical discount factor than the initially maximal one. We formalize this argument in the Appendix, accounting for the potential discontinuity of $B(\cdot|\delta)$ in δ.

[12] $B(\cdot \mid \delta)$ is not continuous with respect to δ because optimal experimentation is not.

In contrast, for small values of m, an initially limited experimentation policy may or may not be easier to implement than an initially maximal one. As shown in Proposition 4, exploration and exploitation decisions are optimally co-determined across domains. Thus, delaying exploration in domain j not only reduces its associated surplus but may also lower the surplus in all other domains. In the Appendix, we show that, when m is small, this advantage of initially maximal policies can outweigh the benefits of initially limited policies discussed above. To demonstrate this, we construct a distribution of project values that yields significant advantages of conducting multiple explorations in parallel, making initially limited policies suboptimal for all discount factors. These intuitions are consolidated in the following proposition.

Proposition 5 (**Initial Scope of Experimentation**)
Suppose $(1-\alpha)\mathbb{E}(v_p) < c$ *and* $m > 1$*. Two thresholds* $\delta^* \leq \bar{\delta} < \delta^0$ *exist such that:*

1. *If* $\delta \geq \bar{\delta}$*, any optimal relational contract is such that the scope of experimentation is initially maximal.*
2. *If* $\delta \in [\delta^*, \bar{\delta})$*, any optimal relational contract is such that the scope of experimentation is initially limited.*
3. *If* $\delta < \delta^*$*, the scope of experimentation is equal to zero in all periods.*

Further, denote $m^* := \sup_{m \geq 2}\{m : \bar{\delta} = \delta^*\}$*. An initially limited experimentation policy is optimal for intermediate discount factors for large* m *(i.e.,* $m^* < \infty$*), but may never be optimal for small* m *(i.e.,* $m^* > 2$ *may occur).*

The previous proposition established results on the players' initial scope of experimentation but did not address its long-term dynamics. We now present findings on their terminal scope. Any non-empty experimentation policy—whether initially limited or initially maximal—has a strictly positive probability of becoming terminally maximal, as players may always, by chance, identify a project valuable enough to sustain the first-best policy indefinitely. Moreover, as discussed above, the optimality of initially limited policies relies crucially on the prospect of sufficiently likely subsequent scope expansions. These observations raise a broader question: is experimentation scope guaranteed to be maximal—and therefore efficient— asymptotically?

Proposition 6 (**Terminal Scope of Experimentation**)
The following statements hold:

1. *There exist optimal experimentation policies that are both initially limited and, with strictly positive probability, terminally limited.*

2. *There exist optimal experimentation policies that are both initially maximal and, with strictly positive probability, terminally limited.*

The reason why the players' scope of experimentation may be terminally limited on path can be understood by considering a vector of project values, $\hat{\mathbf{v}}$, and a subset of domains $s \subset \{1, \ldots, m\}$ for which:

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

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

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

To prove the first statement of the proposition (respectively, the second statement), in the Appendix we show that a), b), and c) hold simultaneously under an initially limited (respectively, initially maximal) policy. Intuitively, these dynamics arise only when $\hat{\mathbf{v}}$ is high enough for a) and c) to hold but low enough for b) to be satisfied.[13]

In this subsection, we analyzed the dynamics of the players' scope under second-best experimentation. We showed that for intermediate discount factors and large maximal potential scope m, the players find it optimal to begin with limited scope, an approach made credible by the possibility of many subsequent scope expansions

[13]The fact that terminally limited scope may arise for intermediate values of $\hat{\mathbf{v}}$—and consequently for intermediate values of the relationship—is reminiscent of Chassang (2010)'s result, where exploration may cease when some but not all productive actions have been "revealed," leaving the value of the relationship in an intermediate range. Despite the differences in setting, the core intuition is similar: conducting additional exploration requires halting the exploitation of an existing project. In our setting, the newly explored project cannot be exploited in the current period due to b). In Chassang's setting, the absence of transferable utility means that exploring an additional action may require terminating the relationship, thereby sacrificing some future exploitation. The difficulty in computing the endogenous loss from these forgone exploitations in closed form is precisely what hinders analytical characterizations in both settings.

created by the discovery of valuable projects in the early domains of cooperation. Because the discovery of such projects is path-dependent, the players may end with a permanently limited and thus inefficient scope of experimentation.

5 Further Analysis and Extensions

This section extends our analysis in three directions. First, we examine how the maximum potential scope of experimentation influences its feasibility and profitability. Second, we analyze a concrete example to graphically illustrate some of the key dynamics of the model. Finally, we explore several simple extensions in which the domains of cooperation are not identical or independent.

5.1 Comparative Statics of Scope

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

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

Let $\tilde{\pi}(m) := \pi(m)/m$ denote the average joint surplus per domain of the collaboration. Recall that $\delta^*(m)$ represents the minimum discount factor for which the optimal relational contract is non-empty. For a scaling factor $k \geq 1$, the following weak inequalities follow from Bernheim and Whinston (1990): $\tilde{\pi}(mk) \geq \tilde{\pi}(m)$ and $\delta^*(mk) \leq \delta^*(m)$.[14] In our setting, we can provide necessary and sufficient conditions

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

for these inequalities to hold strictly, due to the dynamics stemming from the players' exploration of projects. Specifically, $0 < \delta^*(m\cdot k) < \delta^*(m)$ for $k > 1$ if $(1-\alpha)\mathbb{E}(v_p) < c$ and otherwise $\delta^*(m \cdot k) = 0$ regardless of k. When $(1-\alpha)\mathbb{E}(v_p) < c$, the optimal relational contract will be empty for low discount factors. In these instances, scaling up m will strictly decrease δ^*. To see why, note that if the players were to implement k independent and concurrent collaborations, each with an identical experimentation policy, the threshold $\delta^*(m \cdot k)$ would be independent of k. However, this approach would be inefficient as it only leverages relational interdependencies within segmented multi-domain experimentation policies. Therefore, the players could sustain a non-empty relational contract for lower discount factors by leveraging interdependencies across all $m \cdot k$ domains. By an identical reasoning, $\tilde{\pi}(m \cdot k) > \tilde{\pi}(m)$ whenever the second-best experimentation policy is non-empty.

5.2 Multi-Project Collaborations: A Graphical Illustration

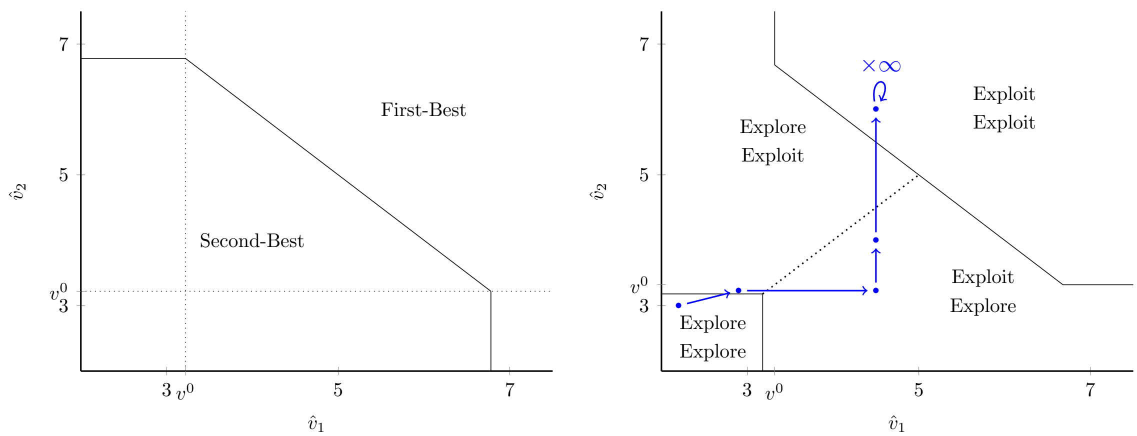
We analyze an example with specific parameter values. We set $c = 1$ and $\delta = 1/3$. Furthermore, we consider a symmetric relationship by setting $\alpha = 1/2$. The players can cooperate in two domains ($m = 2$). Finally, the project values v_p are drawn from a shifted exponential distribution with 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 (5) does not hold, players either: (i) exploit one project while exploring another, or (ii) explore two projects simultaneously.

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

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

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

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.[15]

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

5.3 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 a two-domain version of our framework by supposing that that the first domain is exactly as in the main model, while the other contains a single project with either low or high value. We show that even when

[15] 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.

immediate cooperation across both domains is feasible, players may choose to postpone exploring the risky domain 2 project. This delay continues until a sufficiently valuable project is discovered in domain 1. Such a gradual approach safeguards the collaboration against complete dissolution should the radical innovation fail.

Can "win-win" projects serve as stepping-stones?
In the main analysis, we assumed that each project's benefits accrue to only one player. However, the model can be extended to reflect more nuanced real-world scenarios. Collaborating parties often engage in both "win-win" projects yielding mutual benefits and projects that disproportionately advantage certain participants. In modeling these scenarios, this extension assumes two domains with distinct benefit structures. In one domain, projects yield equal benefits to both players.[16] The other domain follows the main analysis, where project benefits accrue exclusively to one player. We show that optimal experimentation is initially limited for low values of the discount factor and that the domain with symmetric projects is explored first.

How do technological interdependencies influence gradualism?
In the third extension, we introduce positive correlation between project values across domains, such that discovering a valuable project in one domain immediately reveals a project of equal value in the other. This assumption reflects how success in one area can enhance opportunities in another (e.g., mRNA technology's wide applicability across medical conditions). Absent incentive issues, players would optimally explore both domains concurrently to expedite valuable project discovery. With asymmetric benefits, an initially limited approach is strictly optimal for intermediate discount factors. These findings suggest initially limited approaches are more likely to be optimal in R&D environments with stronger cross-domain knowledge spillovers.

6 Applied Insights

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

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

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. 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) highlight the dynamic nature of buyer-supplier relationships, emphasizing the central role of relational contracts. They describe an initial "search and trial phase" that evolves into an "expansion phase," characterized by increased risk-taking and deeper mutual dependence. As they note, "The rudiments of trust and joint satisfactions established in the exploration stage now lead to increased risk taking within the dyad. Consequently, the range and depth of mutual dependence increase." A senior executive from a Toyota supplier similarly described their relationship with Toyota: "We started by making one component, and as we improved, [Toyota] rewarded us with orders for more components" (Liker and Choi, 2004). The common pattern of these relationships starting small before expanding is consistent with our findings, particularly Proposition 5, which shows the potential optimality of gradual expansion in collaborative scope. It also supports our extension in Section 5.1, which examines 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 6, showed that the long-term scope of a collaboration is determined during the initial phases, with early outcomes 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).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 a complementary rationalization of explanation (ii) by showing that the long-run scope of collaboration may be inefficiently limited. When players transition from exploration to exploitation in one domain, they may lose the ability to cooperate in other domains, potentially resulting in limited scope.

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

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, 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. Third, 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.

While our primary focus is on buyer-supplier dynamics and firm-level productivity, our framework extends to political economy settings involving multi-domain collaboration. In federal systems, central governments use fiscal transfers to incentivize subnational policy experimentation (c.f. Callander and Harstad, 2015; Wang and Yang, forthcoming). Our analysis highlights substantial path dependence in policy implementation and suggests a priori unexpected spillovers across policy domains. Moreover, political and economic unions like the European Union facilitate cross-domain collaboration through shared resources and structural funds, with members retaining the option to exit. Consistent with our analysis, the formation of the EU involved gradual step-by-step integration, initially prioritizing mutually beneficial projects before expanding to more ambitious policies with unevenly distributed costs and benefits (see, e.g., Spolaore, 2015, and references therein, as well as Section 5).

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gibbons, 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.

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

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

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

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

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

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

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

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

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," American Economic Review: Insights 2025. Forthcoming.

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

Proof of Proposition 1. Following a reasoning almost identical to that in Bergemann and Välimäki (2001), player 0 treats each domain independently and identically and never recalls a project because $|\mathcal{P}_j| = \infty \; \forall j$. Therefore, the optimal policy conditions only on the project with the highest value amongst all previously explored projects, whose value we denote $\hat{v}$. The Bellman Equation for player 0 is:

$$B^0(\hat{v}) = \max_{\text{explore, exploit } \hat{v}} \left\{ \mathbb{E}\left(v'\right) - 2c + \delta \mathbb{E}\left(B^0\left(\max\left(\hat{v}, v'\right)\right)\right), \hat{v} - 2c + \delta B^0\left(\hat{v}\right) \right\}. \quad (10)$$

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

$$\frac{1}{1-\delta}(v^0 - 2c) = \mathbb{E}(v_p - 2c) + \frac{\delta}{1-\delta}\mathbb{E}(\max\{v, v^0\} - 2c), \tag{11}$$

where standard comparative statics arguments imply that v^0 is increasing in δ. $\square$

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

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

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

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

$$\mathcal{C}_1^2 + b_{2,1}^2 - b_{1,2}^2 > \mathcal{C}_1^1 + b_{2,1}^1 - b_{1,2}^1, \tag{12}$$
$$\mathcal{C}_2^2 + b_{1,2}^2 - b_{2,1}^2 = \mathcal{C}_2^1 + b_{1,2}^1 - b_{2,1}^1. \tag{13}$$

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

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

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

$$\sum_{p \in \hat{\mathbf{P}}^t} \sum_{i=1,2} \max\Big(0, c - \mathbb{E}(x_p \cdot v_p | h^t)\Big) \leq \mathcal{C}(h^t), \tag{14}$$

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

To show that (14) is a necessary and sufficient condition, consider a set of transfers $b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \geq 0$ to be paid on path given a vector of realized values $\mathbf{x}^t, \mathbf{v}^t$.

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

$$\sum_{p\in \mathbf{P}^t} \max\big(c - \mathbb{E}(x_p v_p | h^t), 0\big) \leq \mathbb{E}\Big(b_{-i,i}(\mathbf{x}^t, \mathbf{v}^t) - b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \\ + \sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)\Big), \quad \forall i, \tag{15}$$

$$b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) \leq \sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t\mathbf{x}^t, \sqcup \mathbf{v}^t), \quad \forall \mathbf{v}^t, \forall i. \tag{16}$$

Expectations are taken over the project valuations realizations $\mathbf{x}^t, \mathbf{v}^t$ and $h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t$ denotes the players' updated beliefs after observing $\mathbf{x}^t, \mathbf{v}^t$.[17] The first expression states that the promised transfers and the expected share of the total continuation value must be enough to prevent a player from shirking on any subset of the projects. The second expression states that the each player is willing to pay the other player the necessary transfer.

To show necessity: Note that since Equation (15) must hold for a fixed i, the inequality also holds summing over all i. Further, all transfers cancel out when summing over i. Finally, by definition, $\mathbb{E}(\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)) = \mathcal{C}(h^t)$. Hence, we are left with Equation (14).

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

SubStep 1: We show it is necessary and sufficient to replace Equation (16) by its expectation. This new expression is as follows:

$$\mathbb{E}(b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)) \leq \mathbb{E}\Big(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t)\Big) \quad \forall i. \tag{17}$$

We first show that if there is a solution to Equations (17) and (15), then there exists a solution to Equations (16) and (15).

Take a set of transfers $b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)$ that satisfy Equations (17) and (15). Define:

$$b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t) = \sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t) \\ - \Bigg(\mathbb{E}\Big(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t) - b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)\Big)\Bigg). \tag{18}$$

Since Equation (17) holds, the term in the expectation of Equation (18) is positive

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

and thus Equation (16) holds for all realizations of $\mathbf{x}^t, \mathbf{v}^t$ under the set of transfers $b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)$. Finally, $\mathbb{E}(b'_{i,-i}(\mathbf{x}^t, \mathbf{v}^t)) = \mathbb{E}(b_{i,-i}(\mathbf{x}^t, \mathbf{v}^t))$ so Equation (17) continues to hold.

SubStep 2: Using substep 1, it suffices to show that Equation (14) implies a solution to Equations (15) and (17). To simplify all the notation with expectations, Equation (15) can be re-expressed as:

$$\beta_i - \gamma_i \leq (\tilde{b}_{-i,i} - \tilde{b}_{i,-i}), \tag{19}$$

where $\tilde{b}_{i,-i}$ is the expected transfer from i to $-i$, $\beta_i = \sum_{p \in \mathbf{P}^t} \max\big(0, c - \mathbb{E}(x_p v_p | h^t)\big)$, and $\gamma_i = \mathbb{E}(\sigma_i(\mathbf{x}^t, \mathbf{v}^t)\mathcal{C}(h^t \sqcup \mathbf{x}^t \sqcup \mathbf{v}^t))$. Equation (17) can thus be re-written as:

$$\tilde{b}_{i,-i} \leq \gamma_i. \tag{20}$$

Rearranging Equation (14) implies $\sum_i (\beta_i - \gamma_i) \leq 0$. One can now show that $\tilde{b}_{i,-i} = \max(0, \beta_{-i} - \gamma_{-i})$ satisfies Equation (20). Further, Equation (19) holds because:

$$\beta_i - \gamma_i \leq \max(0, \beta_i - \gamma_i) - \max(0, \beta_{-i} - \gamma_{-i}) \tag{21}$$

$$\iff \max(0, \beta_{-i} - \gamma_{-i}) - \min(0, \gamma_i - \beta_i) \leq 0 \tag{22}$$

$$\Longleftarrow \sum_i (\beta_i - \gamma_i) \leq 0, \tag{23}$$

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

Finally, Equation (20) reduces to

$$\max(0, \beta_{-i} - \gamma_{-i}) \leq \gamma_i \Longleftarrow \beta_{-i} - \gamma_{-i} \leq \gamma_i \tag{24}$$

$$\Longleftarrow \sum_i (\beta_i - \gamma_i) \leq 0, \tag{25}$$

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

Step 3: We show that any relational contract that implements a given experimentation policy can be replaced by an alternative relational contract that implements the same experimentation policy and yields no surplus to player 2. First, note that the way the players share their continuation value does not affect Equation (2) from the main text. Hence, for any period t where player 2's expected payoff is positive,

$w_{2,1}$ can be increased until player 2's expected payoff is zero. Player 2 is willing to make this transfer because not doing so would be seen as a deviation, resulting in a payoff of 0 for player 2.

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

Proof of Proposition 3. When the players have identified projects with values $\hat{v}_1, \dots, \hat{v}_m$ 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 $\hat{v}'_1, \dots, \hat{v}'_m$, the players exploit $\hat{v}'_j$ if and only if $\hat{v}'_j \geq v^0$. This condition is as follows:

$$c \sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j \geq v^0} + \max\{0, c - (1-\alpha)\mathbb{E}(v_p)\} \sum_{j=1}^{m} \mathbb{1}_{\hat{v}'_j < v^0} \leq \sum_{j=1}^{m} \mathcal{C}^0(\hat{v}'_j), \tag{26}$$

$\forall (\hat{v}'_1, \dots, \hat{v}'_m) \geq (\hat{v}_1, \dots, \hat{v}_m)$, which corresponds to (2) when the players implement the first-best policy and where $\mathcal{C}^0(\hat{v}'_j)$ denotes the continuation value associated with domain j under the first-best policy. Note that $\mathcal{C}^0(\hat{v}'_j)$ (i) is constant below v^0, (ii) is such that $\lim_{x \uparrow v^0} \mathcal{C}^0(x) > \lim_{x \downarrow v^0} \mathcal{C}^0(x)$ and (iii) is increasing above v^0. Given such properties, setting $\hat{v}'_j = \max\{\hat{v}_j, v^0\}$ both minimizes the right-hand side and maximizes the left-hand side of (26). Thus, an equivalent condition is:

$$m \cdot c \leq \delta \Big(\sum_{j=1}^{m} \frac{1}{1-\delta} (\max\{\hat{v}_j, v^0\} - 2c) \Big). \tag{27}$$

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

Proof of Corollary 2. The characterization of $v^*(\delta)$ follows from Corollary 1. The existence of δ^* follows an identical argument to that made in Proposition 5. Suppose $\delta < \delta^0$ and consider the policy described in the corollary. By definition of $\tilde{v}$, it suffices to check that the policy is implementable in period 1 (i.e., satisfies Inequality (4)), which is most binding when $\alpha = 1$ (henceforth assumed). At $\delta = \delta^0$, this constraint is as follows: $c \leq \mathcal{C}^0(\text{explore})$, where $\mathcal{C}^0(\text{explore})$ is defined by:

$$\frac{v^0 - 2c}{1-\delta} = \mathbb{E}(v_p - 2c) + \mathcal{C}^0(\text{explore}) \implies \mathcal{C}^0(\text{explore}) > \frac{\delta}{1-\delta}(v^0 - 2c). \tag{28}$$

δ^0 is defined by: $c = \frac{\delta}{1-\delta}(v^0 - 2c)$. Therefore, for δ slightly below δ^0, the policy in period 1 satisfies Inequality (4) if and only if the continuation value is continuous at δ^0, which holds since $\tilde{v}$ is continuous with respect to δ. □

Proof of Proposition 4. Denote $t(p) = \inf_t\{t : p \in \mathbf{P}^t\}$. By contradiction, $\forall p \in \mathcal{P}$, either (i) $p \in \mathbf{P}^t\ \forall\, t > t(p)$ or (ii) $p \notin \mathbf{P}^t\ \forall\, t > t(p)$. Further, by monotonicity, for each domain j, there exists a threshold $v_j^*(\hat{\mathbf{v}}_{-j})$ such that the players exploit a project with value $\hat{v}_j$ if and only if $\hat{v}_j \geq v_j^*(\hat{\mathbf{v}}_{-j})$, where $\hat{\mathbf{v}}_{-j}$ denotes the values of the best projects found in the remaining domains.

Note that $v_j^*(\cdot)$ is weakly increasing in each of its arguments; otherwise, with positive probability, statement 2 of the proposition would be satisfied. Further, for a sufficiently large $\hat{\mathbf{v}}_{-j}$, the first-best experimentation policy is implementable (Proposition 3), implying that $v_j^*(\hat{\mathbf{v}}_{-j}) \leq v^0$. Therefore, with positive probability, the players permanently exploit only projects with value weakly less than v^0, which would imply (i) $v^0 \geq \tilde{v}$ and, thus, (ii) that $\delta \geq \delta^0$. □

The following lemma will aid in proving the next two propositions. Let F denote a distribution of v_p with finite support. Consider a sequence of continuous approximations F_n such that $F_n \leq_{\text{F.O.S.D}} F_{n-1}$, $F_n \geq_{\text{F.O.S.D}} F\ \forall n$, and $F_n \to F$. Define the optimal experimentation policy $\mathbf{a}(\cdot)$ as *strict* if the following conditions hold for all $\hat{\mathbf{v}}$: (i) $\mathbf{a}(\hat{\mathbf{v}})$ satisfies Inequality (4) strictly, (ii) if $\mathbf{a}'(\hat{\mathbf{v}})$ is preferred to $\mathbf{a}(\hat{\mathbf{v}})$, then $\mathbf{a}'(\hat{\mathbf{v}})$ fails Inequality (4) strictly, and (iii) if $\mathbf{a}'(\hat{\mathbf{v}})$ satisfies Inequality (4), the players strictly prefer $\mathbf{a}(\hat{\mathbf{v}})$ over $\mathbf{a}'(\hat{\mathbf{v}})$. Let $B^n(\cdot)$ denote the associated Bellman equation with F^n, and $\mathbf{a}^n(\cdot)$ the corresponding optimal experimentation policy.

Lemma 1 (**Discretization**)
For any $\hat{\mathbf{v}} \in \operatorname{supp} F^m$*, if the optimal experimentation policy is strict, then (i)* $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$ *and (ii) for all* $\hat{\mathbf{v}} \in \operatorname{supp} F^m$*,* $\mathbf{a}^n(\hat{\mathbf{v}}) \to \mathbf{a}(\hat{\mathbf{v}})$*.*

Proof of Lemma 1. First note that (i) $\implies$ (ii). By contradiction, suppose $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$ for all $\hat{\mathbf{v}}$ but there exists a $\hat{\mathbf{v}}^*$ such that $\mathbf{a}^n(\hat{\mathbf{v}}^*) \not\to \mathbf{a}(\hat{\mathbf{v}}^*)$. As $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$ for all $\hat{\mathbf{v}}$, then $\mathcal{C}^n(\hat{\mathbf{v}}^*, \mathbf{a}) \to \mathcal{C}(\hat{\mathbf{v}}^*, \mathbf{a})$ uniformly with respect to $\mathbf{a}$ (as $\mathbf{a}$ belongs to a finite set). However, given that the preference at $\hat{\mathbf{v}}^*$ is strict, we must have $\mathbf{a}^n(\hat{\mathbf{v}}) \to \mathbf{a}(\hat{\mathbf{v}})$, a contradiction.

Let us now prove (i) by contradiction. Note that $\hat{\mathbf{v}}$ has a lattice structure. If (i) fails, there exists a $\hat{\mathbf{v}}^*$ such that $B^n(\hat{\mathbf{v}}^*) \not\to B(\hat{\mathbf{v}}^*)$ but $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$ for any $\hat{\mathbf{v}} > \hat{\mathbf{v}}^*$.

Claim 1: Denote by $\bar{v} = \sup \operatorname{Support}\{v_p\}$, then $\hat{\mathbf{v}}^*$ cannot correspond to $\bar{v}, \dots, \bar{v}$. Note that $\bar{v} > v^0$. Therefore, there exists an n^* for which, if $n > n^*$, the players permanently exploit in all domains when $\hat{\mathbf{v}} = \bar{v}, \dots, \bar{v}$. Claim 1 follows by noting that the net-present value of this policy is identical under $B^n(\cdot)$ and $B(\cdot)$.

Let us now consider $\hat{\mathbf{v}}^* \neq \bar{v}, \dots, \bar{v}$. By assumption, $\forall \hat{\mathbf{v}} > \hat{\mathbf{v}}^*$ $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$. Further, as $B^n(\cdot)$ is decreasing and bounded below by $B(\cdot)$, $B^n(\hat{\mathbf{v}}^*) \not\to B(\hat{\mathbf{v}}^*) \implies B(\hat{\mathbf{v}}^*) < \lim B^n(\hat{\mathbf{v}}^*)$. Next, as $B^n(\hat{\mathbf{v}}) \to B(\hat{\mathbf{v}})$ $\forall \hat{\mathbf{v}} > \hat{\mathbf{v}}^*$, then $\mathcal{C}^n(\hat{\mathbf{v}}^*, \mathbf{a}) \to \mathcal{C}(\hat{\mathbf{v}}^*, \mathbf{a})$. As $B(\hat{\mathbf{v}}^*) < \lim B^n(\hat{\mathbf{v}}^*)$, then $\lim \mathbf{a}^n(\hat{\mathbf{v}}^*) \neq \mathbf{a}(\hat{\mathbf{v}}^*)$. Given that the continuation value converges, $\mathbf{a}^n(\hat{\mathbf{v}}^*)$ must converge, implying $\mathbf{a}(\hat{\mathbf{v}}^*) \neq \mathbf{a}'(\hat{\mathbf{v}}^*) := \lim \mathbf{a}^n(\hat{\mathbf{v}})$. Therefore, $\mathbf{a}'(\hat{\mathbf{v}}^*)$ must be strictly preferred to $\mathbf{a}(\hat{\mathbf{v}}^*)$. Given our definition of "strict," $\mathbf{a}'(\hat{\mathbf{v}}^*)$ must strictly fail Inequality (3) for $B(\hat{\mathbf{v}}^*)$. However, this leads to a contradiction because the continuation value has been proven to converge, implying that $\mathbf{a}'(\hat{\mathbf{v}}^*)$ cannot satisfy Inequality (3) as $n \to \infty$. □

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

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

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

$$m(c - (1-\alpha)\mathbb{E}(v_p)) \leq B\left(\mathbf{0}|\bar{\delta}(m)\right) - m\mathbb{E}\left(v_p - 2c\right). \tag{29}$$

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

$$c - (1-\alpha)\mathbb{E}(v_p) \geq \bar{\delta}(m)\Pr(v_p > \tilde{v})B(\mathbf{0}|\bar{\delta}(m)). \tag{30}$$

Combining (29) and (30) implies:

$$\begin{aligned} &\bar{\delta}(m)\Pr(v_p > \tilde{v})B(\mathbf{0}|\bar{\delta}(m)) \leq \frac{B(\mathbf{0}|\bar{\delta}(m))}{m} - \mathbb{E}(v_p - 2c) \\ \iff\ & B(\mathbf{0}|\bar{\delta}(m))\left(\bar{\delta}(m)\Pr(v_p > \tilde{v}) - \frac{1}{m}\right) \leq -\mathbb{E}(v_p - 2c). \end{aligned} \tag{31}$$

However, as shown in the text $B(\mathbf{0}|\bar{\delta}(m))$ diverges to infinity. Hence, $\bar{\delta}(m)\Pr(v_p > \tilde{v}) - \frac{1}{m}$ converges to zero. However (i) $\bar{\delta}(m)$ remains bounded away from zero when $\mathbb{E}(v_p)(1-\alpha) < c$, which has been assumed, and (ii) if $\bar{\delta}(m)$ remains bounded away from zero, then $\tilde{v}$ remains bounded above, implying $\Pr(v_p > \tilde{v})$ remains bounded away from zero, proving that $\bar{\delta}^*(m) < \bar{\delta}(m)$ when m is large.

Finally, we consider a discrete support distribution and leverage Lemma 1 to show that $\delta^* = \bar{\delta}$. We consider a three-point support of benefits: $\{0, \underline{v}, \bar{v}\}$. Further, the experimentation policy described below will be shown to satisfy the definition of "strict" employed in Lemma 1. We also assume $\mathbb{E}(v_p) = 2c$, $\alpha = 1$, and $m = 2$ when listing sufficient inequalities for $\delta^* = \bar{\delta}$ to hold. The following inequalities ensure that there exists a δ such that $|\mathbf{P}^1| = 1$ is not feasible but $|\mathbf{P}^1| = 2$ is:

$$\frac{\underline{v} - 2c}{1-\delta} > \frac{\delta \Pr(\bar{v})}{1 - \delta(1 - \Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} \tag{32}$$

$$2c < \frac{\delta}{1-\delta}(\bar{v} - 2c) \tag{33}$$

$$c > \frac{\delta}{1-\delta}(\underline{v} - 2c) \tag{34}$$

$$2c > \frac{\delta}{1-\delta}(\underline{v} - 2c) + \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} \tag{35}$$

$$c > \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}(\bar{v} - 2c)\frac{1}{1-\delta} + \frac{\delta \Pr(\bar{v})}{1-\delta(1-\Pr(\bar{v}))}\text{npv}^0 \tag{36}$$

$$\begin{aligned} 2c < {} & \frac{2\Pr(\bar{v})(1-\Pr(\bar{v}))}{1-\delta\big(1-2\Pr(\bar{v})(1-\Pr(\bar{v}))\big)}\Big((\bar{v} - 2c)\frac{\delta}{1-\delta} + \text{npv}^0\Big) \\ & + \frac{\Pr(\bar{v})^2}{1-\delta\big(1-\Pr(\bar{v})^2\big)}\frac{\delta}{1-\delta}2(\bar{v} - 2c) \end{aligned} \tag{37}$$

$$\text{npv}^0 = \frac{\delta}{1-\delta}\frac{\Big(\Pr(\bar{v}) + \Pr(\underline{v})\Big)\Big(\frac{\Pr(\bar{v})\bar{v}}{\Pr(\bar{v})+\Pr(\underline{v})} + \frac{\Pr(\underline{v})\underline{v}}{\Pr(\bar{v})+\Pr(\underline{v})} - 2c\Big)}{1-\delta(1-\Pr(\bar{v}) - \Pr(\underline{v}))} \tag{38}$$

Inequality (32) ensures that $v^0 \leq \underline{v}$. Inequality (33) ensures that $\bar{v}, 0$ satisfies Inequality (5). Inequality (34) 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 (35) implies that the players cannot exploit a project worth $\underline{v}$ while exploring in the additional domain. Inequality (36) 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.[18] 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 (38). Finally, Inequality (37) is a necessary condition for the initially 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. We use Mathematica to show that these constraints jointly hold strictly.[19] □

Proof of Proposition 6. We first show that our notion of terminal scope is well defined. If there exists a period t for which the players conduct no explorations, then $\mathbf{P}^{t'} = \mathbf{P}^t$ for all $t' \geq t$. By contradiction, if there exists an equilibrium path where $\liminf |\mathbf{P}^t| <$

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

[19] Code available upon request.

$\limsup |\mathbf{P}^t|$, the players must explore at least one project in each period t. However, for each exploration, with positive probability the players discover a project with value exceeding $m\tilde{v}$, implying that the first-best policy is implementable in all subsequent periods. As a result, $\liminf |\mathbf{P}^t| = \limsup |\mathbf{P}^t|$. This argument also shows that terminal scope equals m with positive probability.

Statement 1: We consider $v_p \in \{0, \underline{v}, \bar{v}\}$, $m = 2$, and $\alpha = 1$. Further, the experimentation policy outlined below will satisfy the definition of "strictness" employed in Lemma 1, implying these results will hold true for continuous approximations. These inequalities ensure the existence of a feasible experimentation policy where the scope of experimentation reaches its maximum of 2 with interior probability, while ensuring no other feasible policy yields a higher joint surplus. We list all the inequalities and comment on each one separately below.

$$2c < \frac{\delta}{1-\delta}(\bar{v} - 2c) + \mathcal{C}^0(\text{explore}) \tag{39}$$

$$c < \frac{\delta}{1-\delta}(\underline{v} - 2c) \tag{40}$$

$$2c > \frac{\delta}{1-\delta}(\underline{v} - 2c) + \mathcal{C}^0(\text{explore}) \tag{41}$$

$$2c > \mathcal{C}^0(\text{explore}) + \frac{\delta}{1-\delta}\Big(\Pr(\bar{v})\bar{v} + (1 - \Pr(\bar{v}))\underline{v} - 2c\Big) \tag{42}$$

$$\begin{aligned}\frac{\underline{v} - 2c}{1-\delta} > v := \mathbb{E}(v_p - 2c) + \frac{\delta}{1-\delta}\Big(\Pr(\bar{v})(\bar{v} - 2c) + \Pr(\underline{v})(\underline{v} - 2c)\Big) \\ + \Big(\Pr(\bar{v}) + \Pr(\underline{v})\Big)\frac{\delta}{1-\delta}(\underline{v} - 2c) + \Big(1 - \Pr(\bar{v}) + \Pr(\underline{v})\Big)\delta v\end{aligned} \tag{43}$$

$$c \leq \mathcal{C}^0(\text{explore}) + \frac{\delta \Pr(\bar{v})\mathcal{C}^0(\text{explore})}{1 - \delta(1 - \Pr(\bar{v}))} \tag{44}$$

Inequality (39) implies that $\{\bar{v}, 0\}$ satisfy Equation (5), where $\mathcal{C}^0(\text{explore})$ was defined in Equation (28). Inequality (40) ensures that the players are able to exploit a project worth $\underline{v}$ in isolation. Inequality (41) ensures that $|\mathbf{P}^t| < 2$ while exploiting the project worth $\underline{v}$ (when the best project found so far on the other domain has value 0). This inequality uses $\mathcal{C}^0(\text{explore})$ as an upper-bound. These statements imply that if the players ever reach a point with a project worth $\underline{v}$, they either exploit the project, explore a project on the other domain while maintaining a scope of 1, or conduct 2 explorations. Inequality (42) ensures that conducting two explorations is not feasible because the upper-bounds associated with the continuation value for the new domain and the domain with a project with value $\underline{v}$ is provided by the first-best policy. Next, Inequality (43) ensures that the players prefer to exploit the project worth $\underline{v}$ as opposed to exploring the domain where the best project is worth 0 until Equation (5) holds and then subsequently implementing the first-best policy. These constraints imply that $|\mathbf{P}^t| = 1$ if the best projects are worth $\underline{v}, 0$. Finally, Inequality (44) ensures that this experimentation policy is feasible. One can check that these constraints, along with (i) $\mathbb{E}(v_p) \geq 2c$ and (ii) $v^0 \leq \underline{v}$, hold jointly.[20]

Statement 2: We prove the result for a discrete support distribution, as Lemma 1 can be used to extend the result to a convex support distribution. We consider $m = 3$ and a trinary support distribution, $v_p \in \{0, \underline{v}, \bar{v}\}$, where $0 < \underline{v} < \bar{v}$.[21] Throughout, let $\tilde{c} := c - (1-\alpha)\mathbb{E}(v_p) > 0$. Let $\bar{p}, \underline{p}$ correspond to the probability that $v_p = \bar{v}$, $\underline{v}$, respectively. For the subsequent argument, consider $\bar{p}$ to be arbitrarily small, in a sense we will make precise below. Suppose

$$2c = \frac{\delta}{1-\delta}(\bar{v} + \underline{v} - 4c), \tag{45}$$

implying that the players can jointly exploit projects worth $\bar{v}$ and $\underline{v}$, but could not permanently exploit a project worth $\bar{v}$ and two projects worth $\underline{v}$. As a result, for any $\tilde{c} > 0$, when $\hat{\mathbf{v}} = \bar{v}, \underline{v}, 0$, the players either (i) permanently exploit the two projects and conduct no additional explorations or (ii) implement a policy involving exploration whose associated surplus is bounded below that of exploiting the project with value $\bar{v}$ and exploring in the other two domains until finding a project with value $\bar{v}$ and subsequently implementing the first best. As the payoff bound of (ii) tends to zero as $\bar{p}$ goes to zero, there exists $\bar{p}^* > 0$ such that for $\bar{p} < \bar{p}^*$, the players choose (i). Hence, it suffices to show that Equation (45), $\bar{p} < \bar{p}^*$, and $|P^1| = m$ may jointly hold. Because the first two of these three conditions are independent of α, they are also independent of $\tilde{c}$. Therefore, holding fixed all remaining parameter values, the continuation value at date one of conducting three explorations is bounded below by $\Pr(v_p = \bar{v})\frac{\delta}{1-\delta}(\bar{v} - 2c)$. As a result, if $\tilde{c}$ is sufficiently small, the players' initial scope will be maximal, thereby completing the proof. $\square$

[20] The code can be provided upon request.

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