The Art of Waiting*

Vasundhara Mallick† Ece Teoman‡

July 28, 2025

Abstract

This paper studies delegated project choice without commitment: a principal and an agent have conflicting preferences over which project to implement, and the agent is privately informed about the availability of projects. We consider a dynamic setting in which, until a project is selected, the agent can propose a project, and the principal can accept or reject a proposed project. Importantly, the principal cannot commit to his responses, and cannot implement a project unless it is proposed. In this setting, the agent has an incentive to hold back on proposing projects that the principal favors so that the principal approves a project favored by the agent. Nevertheless, the principal achieves his commitment payoff in an equilibrium of the game in the frequent-offer limit. This high payoff equilibrium showcases the art of waiting and contrasts with Coasian logic: by giving proposer power to the agent, the principal can credibly commit to rejecting his dispreferred projects until later in the game, giving the agent an incentive to propose principal-preferred projects earlier on. We apply these results to the economics of organization. In particular, these results suggest that to curb a manager's *empire building* plans, eliciting proposals from her "bottom-up" might be better than issuing "top-down" commands.

*We are grateful to S. Nageeb Ali, Vijay Krishna, and Rohit Lamba for their invaluable guidance and continuous support. We greatly benefited from suggestions from Yu Awaya, Ian Ball, Kalyan Chatterjee, Nima Haghpanah, Berk Idem, Navin Kartik, Andreas Kleiner, R. Vijay Krishna, Wenhao Li, Joshua Mollner, and Ran Shorrer. We thank the participants of the Pennsylvania Economic Theory, Midwest Theory, and Women in Economic Theory conferences for their insightful comments.

†Department of Economics, European University Institute. Email: vasundhara.mallick@gmail.com

‡Department of Economics, Florida International University. Email: eceteoman@gmail.com.

1 Introduction

This paper considers a principal-agent problem with the following two features: (i) the agent knows what actions or "projects" are feasible and the principal does not, and (ii) the interests of the two parties are not aligned. Such principal-agent problems abound. Consider the interaction between a CEO (principal) of a firm and a manager (agent), who must jointly choose a project. The manager, due to her expertise, may be better informed which projects the firm *can* undertake, but unlike the CEO, she is motivated by empire building.[1] In such cases, the CEO cannot blithely assume that the manager selects projects that optimize shareholder interests. Another example is that of an antitrust authority deciding which mergers to approve: It only wants to approve those mergers that enhance efficiency or consumer welfare, but firms would like to propose only those mergers that increase their own market power. In such settings, what should the principal do?

These issues have been studied in the literature on *project selection* problems, initiated by the seminal work of Armstrong & Vickers (2010) and Nocke & Whinston (2013). The dominant approach has been to model this as a one-shot interaction between the principal and the agent, where the principal *commits* to the set of projects that he would approve. However, this approach has some limitations. Firstly, such commitment may be unrealistic in many settings, for e.g, if the principal and the agent are part of the same organization and interact over several periods. In this case, if the agent responds to the principal's *permission set* by proposing nothing, the principal may infer that such projects are infeasible and allow other projects, which may be more preferred by the agent. Anticipating that the principal will eventually capitulate, the agent may then wish to hold back on proposing projects the principal prefers. The second limitation is that although this approach gives the principal commitment power, there is no opportunity for him to *elicit* the agent's private information, and learn about which projects are feasible over time. His decision is simply based on his prior beliefs about which projects are feasible.

Motivated by these limitations of the static game, we investigate this question in a dynamic framework. There is a principal and an agent who must jointly choose a project to implement. There is a set of *possible* projects, but some of them may

[1] In an organizational context, empire building refers to the agent's attempt to increase her power and influence within the organization, which may lead her to prefer *flashy* projects rather than those that optimize shareholders' interests.

not be feasible. The agent is privately informed about *which* projects are feasible at time 0. In each round $t \in \{0, 1, 2, \ldots\}$, the agent can propose a feasible project or stay silent; if a project is proposed, the principal can either accept or reject it. This process continues until a project is accepted – in which case, players obtain payoffs from that selected project— or no proposed project is *ever* accepted, so all players obtain payoffs from the status quo, which we normalize to zero. We consider the frequent-offer limit of this model, a sequence of games where the period length vanishes.

If the principal can *commit* to his strategy in this game, then this is obviously an improvement over the static game. This is because now the principal can commit to what he would accept, as a *function* of what the agent proposes. So, he can screen different types of the agent more effectively. But what happens in an *equilibrium* of this game, where the principal cannot commit to his responses to future proposals. One would anticipate that in this case, the principal would suffer a significant loss of payoffs compared to if he could commit in the dynamic game, and even compared to the static game. After all, given the logic we sketched earlier, if no principal-preferred projects are proposed for a while, the principal may capitulate and accept a project that he dislikes. The agent, anticipating such capitulation, would not propose principal-preferred projects, but would rather keep proposing projects that she likes.

Therefore, in absence of commitment to future actions, it is not clear how well the principal can do in the dynamic game. Moreover, the principal also lacks *proposal power* here; he can merely accept or reject projects that the agent proposes. Even complete information intuitions suggest that this will result in the loss of bargaining power for the principal and he will do poorly in the dynamic game. However, our main result, informally stated, is

> **Main Result**. *In the frequent-offer limit, the principal attains his commitment payoff in an equilibrium of the game.*

There is always an equilibrium of the dynamic game where, if the the principal-preferred project is feasible, the agent proposes it at $t = 0$. It is only if this project is **not** feasible, that she proposes the project that *she* prefers, and the principal dislikes. This high payoff equilibrium stipulates that if the agent *only* has the project that she greatly prefers to other projects, and the principal disprefers (but prefers to the status quo), then she *waits* for a certain number of rounds before proposing it. If this project is proposed earlier than stipulated, then the principal rejects it. We show that this behavior is sequentially rational, even at histories where the principal attributes probability 1 to the agent only having such projects. If the agent proposes such projects earlier than specified, the principal believes that the agent must have other projects at her disposal, and rejects the proposal. Such "punishment through beliefs" incentivizes the agent not to propose such projects earlier than stipulated and as such solves the principal's commitment problem. Because the agent anticipates such delays to get her preferred projects approved, she is willing to propose, as in the commitment benchmark, feasible projects that the principal prefers (and she may disprefer).

The key idea underlying our main result is that endowing the agent with the right to propose, along with restricting the principal's action to accepting or rejecting a proposed project, circumvents the principal's commitment problem. To highlight the importance of the *lack* of proposal power for the principal, when he cannot commit to his future actions, we compare our game with another game studied in the literature, where the principal makes all the offers. In each period, the principal restricts the set of projects that the agent can choose from, but cannot commit to not relaxing these restrictions in future rounds. In this case, as Li (2024) shows, Coasian forces take over, and the principal ends up granting full discretion to the agent at the outset.

What implications do our results have for organizational design, and the kind of authority the principal should have? The results for our game, as well as the above comparison seem to suggest that a somewhat *centralized* structure, where the agent brings proposals to the principal, who has the final authority to approve or reject them, is better than a structure where the principal delegates. This seems to go against the insights from the literature on delegation, which says that the principal might be better off delegating to an agent to benefit from her superior information. However, this is not entirely true. To understand what exactly the results mean, it is important to understand the settings that they are applicable in.

In our setting, i) the agent has private information about which decisions are *feasible*, ii) a decision cannot be implemented without the involvement, or cooperation of the agent, and iii) there is a conflict of interest between the principal and the agent. To see why the second part is true, observe that in our dynamic game, the principal can *only* implement a project if the agent proposes it. He cannot simply go ahead and implement a project on his own.[2] And even in the principal-proposes game, which

[2] We might think that this is because the principal does not know whether any given project is

we compare our game to, the principal can restrict the agent's choices, but the agent needs to choose a project from this set in order for it to be implemented.

What are some real-world settings that have these features? One example is the firm has to implement a project that requires specialized knowledge close to the technological frontier, so only the manager understands exactly what is feasible, and how exactly to execute it. However there is a conflict of interest between the CEO and the manager, because the manager favors projects that are in line with her empire building plans, and the CEO wants to implement a project that maximizes the interests of shareholders. Another example is when the organization is going through a restructuring, so the *project* basically refers to what changes should be implemented across various divisions of the organization. A manager heading a division is better informed about the kind of changes her division would be able to adapt to, and her cooperation is essential in implementing these changes. In a non-organizational context, in the merger approval example we gave earlier, the antitrust authority can only approve or reject a merger that is proposed, it cannot force two firms to merge even if it knows that the merger is feasible.

So, our setting is one where the agent already has a lot of discretion, because of the nature of her private information (which is about feasibility), and because a project cannot be implemented unless she proposes it (or chooses it, in case of the dynamic delegation game). We interpret our results as saying that in such settings, we should try to **limit** the agent's discretion, by having the principal retain the authority to veto or rubber-stamp proposals. From an applied perspective, another way to think about our setting is that we are at an **intermediate** stage where some delegation has already happened in the past. This could be delegation of running a division, or of knowledge or skill acquisition. Our results then suggest that in such a setting, where considerable delegation has already happened in the past, the principal should retain the final authority to approve projects.

Finally, we showed that the key idea behind our main result was the existence of an equilibrium where the principal-preferred project is proposed by the agent right away if its feasible, because the project that the agent prefers (and the principal dislikes), is only accepted if proposed with a delay. How should we interpret this *waiting* or

feasible or not. But we view the *complete information* counterpart of our game, where the set of feasible projects is common knowledge, to *also* be one where the principal has to wait for an agent to propose a project to implement it. Even there, he cannot unilaterally implement a project.

delay in terms of applications of our model? It seems natural that delay could arise in implementing decisions that are not seen to be in the best interests of the organization. This may be especially true if the principal believes that the agent is proposing this decision to advance her own interests within the organization (for example, projects that are flashy, and are better for the agent from an empire building point of view). We interpret this delay as arising from the agent's attempts to convince the principal that what he prefers is not feasible, before ultimately proposing the project the principal dislikes. Due to the technical nature of the project, the principal may not understand much; he might be unable to distinguish between genuine evidence that the good project is not feasible, and deception by the mixed type, if she decides to mimic the bad type. However, in equilibrium, the principal understands the the mixed type would not go down this path because of the delay it involves.

The rest of the paper is organized as follows. We first present the related literature. Section 2 describes the model and our dynamic game. Section 3 first introduces a commitment benchmark, and then presents the main result that says that this commitment payoff can actually be attained in an equilibrium of our dynamic game. It then provides a discussion of off-path beliefs and refinements in subsections 3.2.1 and 3.2.2. Section 3.3 compares our extensive form to another dynamic game to highlight the importance of lack of proposal power for the principal. Section 4 provides sufficient conditions under which commitment payoff can be attained in equilibrium for the general N projects case. Section 5 concludes.

1.1 Related Literature

Our paper studies a project selection problem where the agent privately knows which projects are feasible. This problem has been studied in a static setting by Armstrong & Vickers (2010) and by Nocke & Whinston (2013), in the context of merger review. Armstrong & Vickers (2010) is more closely related to our analysis: They study a model where the principal commits upfront, to which projects he would approve. As we described in the Introduction, in our dynamic game, the principal lacks the ability to commit to his responses to future proposals. Nocke & Whinston (2013) study a similar problem in a static setup, in the context of mergers. An antitrust authority can commit ex ante to its merger-approval rule. However, this is not a direct static counterpart of our setup. There are multiple firms (agents) here and given the set of permitted and feasible mergers, the implemented merger is the result of a bargaining process among firms.

Nocke & Whinston (2010) study a dynamic merger review game where merger proposals occur over time. However, their setup is quite different from ours and they derive conditions under which a myopic merger policy is optimal, whereas our paper highlights how delay can emerge as a signaling device in equilibrium. Closest to our work is the dynamic game studied by Schneider (2015), where the agent can only propose feasible projects and the principal approves or rejects them. However, the agent's private information is *richer* than our game, in the sense that in addition to knowing whether a given project is feasible or not, the agent also privately knows the realization of the project, that would determine the payoff to the principal and the agent if the project is implemented. Like us, they describe a class of *waiting equilibria*, but their results are a bit different because of the richer type space. In particular, unlike our game, the equilibrium with the shortest delay is never optimal. Moreover, unlike us, they do not characterize the optimal commitment payoff, or show that the principal-optimal equilibrium achieves it.

More broadly, our paper is also related to the literature on constrained delegation, starting with Holmström (1984), and continued by Melumad & Shibano (1991) and Alsonso & Matouschek (2008). The similarity to our work is that the principal is at an informational disadvantage relative to the agent. The difference is that in our setting, the principal's disadvantage is not just informational: He cannot implement a project without the agent's cooperation. Moreover, the nature of the private information the agent has is different. In these other papers, the agent's private information is relevant to the payoff from a particular action, but there is, unlike our paper, no certainty about which actions are available. Also related is Dessein (2002), who shows that the best way to make use of a biased agent's superior information could be to delegate the decision to the agent, rather than communicate with her, in absence of commitment power. We, on the other hand, show that the principal can attain his commitment payoff in a game where he elicits information from the agent. Again, the nature of private information, and the fact that ours is a dynamic game leads to the different conclusions.

The separating equilibrium in our model, where delay is used as a signaling device, is also reminiscent of the literature on bargaining where time is a strategic variable, notably Admati & Perry (1987), and Crampton (1992). In these papers, the delay in responding to offers, or making a counter offer, is a strategic variable, and is used to signal the strength of a player's bargaining position. However, both these papers consider alternating offer games. In fact, in these games, if the uninformed party lacked proposal power, and the informed party was making all the proposals, then the informed party would end up capturing all the surplus. In contrast to these, in our paper, time is not a strategic variable, and the time between offers is fixed. Also, the lack of proposal power for the principal is crucial for separation to occur in our model.

2 Model

A principal (he) and an agent (she) jointly choose a project to implement. There are two *possible* projects, denoted by $N = \{g, b\}$. Each project $i \in \{g, b\}$ is characterized by a pair of payoffs $(\alpha_i, \pi_i) \in \mathbb{R}^2_{++}$, where α_i is the agent's payoff from implementing the project, and π_i is the principal's. The players have conflicting preferences over the projects: $\pi_g > \pi_b > 0$ and $\alpha_b > \alpha_g > 0$. We refer to g as the ***good*** project (better for the principal), and b as the ***bad*** project (better for the agent).

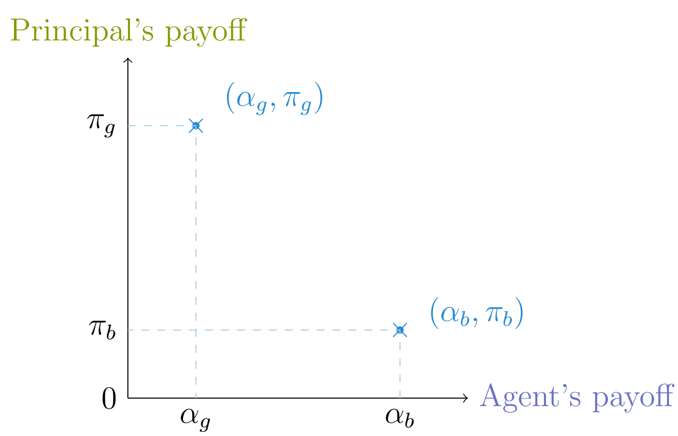

Figure 1: The project space with the principal-preferred good project (α_g, π_g) and the agent-preferred bad project (α_b, π_b).

The challenge is that not every *possible* project may be *feasible*, or *available* to implement, only a subset may be. And which projects are available is the agent's private information, or her *type*. The agent has three possible types:

- $G = \{g\}$, the *good* type with only the good project available;

- $B = \{b\}$, the *bad* type with only the bad project available;

- $M = \{g, b\}$, the *mixed* type with both projects available.[3]

Let $\mathcal{N} = \{G, B, M\}$, the set of all non-empty subsets of the set of possible projects. The set of *feasible* projects, $S \in N$ is drawn according to the prior $\mu_0 \in \Delta(\mathcal{N})$. We assume that $\mu_o(M)$ and $\mu_0(B)$ are both strictly positive. We will see that this is important for a non-trivial conflict of interest between the principal and the agent.

We now describe the dynamic game to be played between the principal and the agent. Time is discrete and the principal and the agent have a common discount factor $\delta \in (0, 1)$. The timing is as follows:

1. Before the beginning of play, the agent's type $S \in \mathcal{N}$ realizes.

2. At each time period $t = 0, 1, 2, \ldots$, the Agent can either (i) **stay silent** (denoted by $\varnothing$), or (ii) **propose a feasible project**, i.e. some $i \in S$.[4]

3. If, at time period t, the agent is silent, there is nothing for the principal to do; both players get a payoff of 0, and the game continues to the next period. If the agent proposes a project i at t, the principal can either (i) **accept**, in which case the **game ends**, and the players get their discounted payoffs, i.e. the principal gets $\delta^t \pi_i$, and the agent gets $\delta^t \alpha_i$ or (ii) **reject**, in which case both players obtain a payoff of 0, and the game continues to the next period. We focus on the case of $\delta \to 1$, which we interpret it as the frequent-offer limit of the game.

We refer to this game as the **dynamic elicitation game**, as it involves the principal eliciting proposals from the agent, which he then accepts or rejects. In this game, for any time period t, the set of all possible histories at the beginning of period t is $\mathcal{H}_t = (N \cup \varnothing)^t$. This captures the fact that if we are at t, for any $t' \leqslant t - 1$, we can have two cases: (i) a project $i \in N$ was proposed and rejected, or (ii) the agent was silent – This is denoted by $\varnothing$. An element of $\mathcal{H}_t$ is denoted by h_t. If the agent is of type S, her strategy, denoted by σ_S, maps any history to a probability

[3] We have implicitly assumed here that at least one project is always feasible. We can also envision an *empty* type E that has no feasible projects. No results change if we allow for such a type, other than making some of the notation more cumbersome.

[4] If we allowed for an empty type $S = E$, then this type's only possible action in any time period would be to stay silent.

distribution over $\{i|i \in S\} \cup \varnothing$. So, at any history h_t, and for any $i \in S \cup \varnothing$, $\sigma_S(i|h_t)$ is the probability with which an agent of type S proposes i. The principal's strategy, denoted by σ_P, maps any history, and a proposal of project i at that history to a probability of accepting i. So, $\sigma_P(i|h_t)$ is the probability of accepting i at history h_t, where $\sigma_P(\varnothing|h_t) = 0$ for all t, and for all $h_t \in \mathcal{H}_t$.[5] We denote the principal's beliefs at history h_t by $\mu(h_t)$, where $\mu_S(h_t)$ is the probability this belief assigns to type S of the agent. Our equilibrium concept is Perfect Bayesian Equilibrium; both players play sequentially rationally and the principal's beliefs about the agent's type are updated according to Bayes' rule whenever possible.

3 Equilibrium

This section presents our main result. We study how *well* the principal can do in an equilibrium of our game, despite being uninformed, and lacking commitment power. We first present a commitment benchmark. This is to show what is the best the principal can do if he can **commit** to a strategy in the dynamic elicitation game; how effectively he can elicit the agent's private information and screen different types of the agent. Then, our main result shows that this commitment payoff is, in fact, attainable in an *equilibrium* of the game, when the principal has no commitment power. In particular, the separation of the mixed and the bad type is possible in an equilibrium of the game. Finally, we compare our extensive form with another extensive form, to highlight some features of our game that are important for this separation to occur in equilibrium.

3.1 Commitment Benchmark

In our game, the principal has no commitment power. This means that at any history, the decision to accept or reject a proposal must be sequentially rational given the principal's beliefs about the agent's type at that history, and he also cannot commit to his responses to future proposals. In this section, we answer the following question:

> *If the principal could* ***commit*** *to a strategy in the game, what's the highest payoff that he can achieve, given that the agent best responds to the his strategy?*

[5]This captures the fact that if agent is silent, then the principal's only possible action is to reject.

It turns out that the relative likelihood of the mixed and the bad type is a sufficient statistic for determining the principal's *optimal commitment strategy*. Let $\lambda := \frac{\mu_o(M)}{\mu_o(B)}$, the relative likelihood of the mixed and the bad type, and $\lambda^* := \frac{(1-\frac{\alpha_g}{\alpha_b})\pi_b}{(\pi_g-\pi_b)}$. Then, the following is true:

Proposition 1. *1. If $\lambda > \lambda^*$, then:*

- ***Principal's optimal commitment strategy** is to (i) accept g with probability one and b with probability $\frac{\alpha_g}{\alpha_b}$ if proposed at $t = 0$,(ii) reject every proposal after $t = 0$, irrespective of history.*
- ***Agent's best response** is to i) propose g at $t = 0$ whenever g is feasible, i.e if she is of mixed or good type, and ii) propose b at $t = 0$ if she is of bad type.*

2. If $\lambda \leqslant \lambda^*$, then:

- ***Principal's optimal commitment strategy** is to accept both g and b whenever proposed, irrespective of history.*
- ***Agent's best response** is to i) propose g at $t = 0$ if she is of good type, and ii) propose b whenever it is feasible, i.e. if she is of mixed or bad type.*

The proof that the above strategies are optimal for the principal is in the Appendix. We provide some intuition below, but first we argue that the agent's strategies described above are indeed best responses to the principal's commitment strategy in each case.

For the $\lambda \leqslant \lambda^*$ case, this is obvious. Since any project will be accepted by the principal at every history, it is optimal for each type of agent to propose her favorite *available* project at $t = 0$. For the $\lambda > \lambda^*$ case, given the principal's strategy, the agent's **only** shot at getting *anything* accepted is at $t = 0$, since the principal rejects every proposal thereafter. For the good and bad types, the choice is obvious, to propose the *only* project they have. For the mixed type, there is a trade-off: she can either propose g and it will get accepted with probability one, or she can propose b, her preferred project, but it will get accepted with probability strictly lower than one. But the probability of acceptance of the bad project is $\frac{\alpha_g}{\alpha_b}$, so the mixed type is *indifferent* and it is indeed a best response to propose g.

So, in each case, what do we have in terms of on-path implementation of projects? In the $\lambda > \lambda^*$ case, conditional on the agent's type being mixed, g is implemented with probability one, and conditional on his type being bad, b is implemented with probability $\frac{\alpha_g}{\alpha_b} < 1$, and with probability $1 - \frac{\alpha_g}{\alpha_b}$, nothing is implemented. When $\lambda \leqslant \lambda^*$, for each type of the agent, b is implemented with probability one. Comparing the two outcomes, conditional on the agent's type being mixed, the principal is better off in the $\lambda > \lambda^*$ case, because he is able to "extract" the good project from the mixed type. However, conditional on the agent's type being bad, he is better off when $\lambda \leqslant \lambda^*$, because here, b, the *only* project is implemented with probability one, as opposed to the $\lambda > \lambda^*$ case, where it is implemented with a probability strictly lower than one. So there seems to be a trade off between *extracting* the good project from the mixed type, and implementing the bad project with probability one when the agent is of bad type.

To see the reason for this trade off, and the intuition behind why the principal's strategy is optimal in each case, first consider the ***complete information benchmark***, where the principal can **observe** the agent's type, or the set of feasible projects. In this case, if the agent's type is mixed, then it is obviously optimal for the principal to commit to a strategy where he **only** accepts g at any history, and rejects b whenever proposed. So even though the mixed type does have the bad project, she knows the principal will never accept it, so it is optimal for her to simply propose the good project at $t = 0$. If, on the other hand, the agent is of bad type, an optimal strategy for the principal is to accept b *whenever* proposed, since it is the **only** feasible project. So, in this case, b is implemented at $t = 0$.

However, when the principal **cannot** observe the agent's type, there is a trade-off. The principal wants to accept the bad project when it is the *only* feasible project, and he also wants to "extract" the good project from the mixed type. But if the principal accepts the bad project whenever it is proposed, then the mixed type would never propose the good project. So, in order to incentivize the mixed type to propose g, the principal has to *lower* the probability with which he accepts the bad project, if proposed. Therefore extracting g from the mixed type, i.e. implementing g with probability one when the agent's type is mixed must come at a cost: conditional on the agent's type being bad, b, the *only* available project, will be implemented with a probability strictly lower than one.

So when is this cost worth incurring? The answer, as described in the above proposition, depends on the relative likelihood of the mixed and bad types. When $\lambda > \lambda^*$, the likelihood of the mixed type is very high, relative to the bad type. So the expected benefit from implementing g when agent's type is mixed outweighs this cost, and the principal finds it optimal to extract g with probability one from the mixed type. So, he reduces the probability of acceptance of b to $\frac{\alpha_g}{\alpha_b}$, which is *exactly* the probability at which the mixed type is indifferent between proposing g and b. When $\lambda \leqslant \lambda^*$, the cost is higher than the benefit, so the principal finds it optimal to **not** extract g from the mixed type, and b is implemented irrespective of the agent's type.

Finally, observe that in both cases ($\lambda > \lambda^*$ and $\lambda \leqslant \lambda^*$), the agent's best response to the optimal commitment strategy ensures that the play concludes at $t = 0$. After $t = 0$, no project is proposed and implemented on-path. In particular, even in the $\lambda > \lambda^*$ case, where the mixed and bad types are *separated*, the principal does not need multiple periods to achieve this separation when he has commitment power. He can do it simply by using the probability of acceptance, in particular, by randomizing over accepting and rejecting the bad project. We will see that this is in contrast to how separation is achieved in an equilibrium of the game when the principal lacks commitment power.

3.2 Attaining Commitment Payoff in Equilibrium

Now that we have solved for the optimal commitment strategy, we turn our attention back to the game, where the principal cannot commit to his responses to the agent's proposals. Therefore, sequential rationality considerations might limit how effectively he can screen different types of the agent.

To see why, consider the case when $\lambda > \lambda^*$. Recall that here, the principal's optimal commitment strategy involves accepting b with probability $\frac{\alpha_g}{\alpha_b} \in (0, 1)$, when proposed at $t = 0$. Recall also that the agent's best response to this strategy is to propose g at $t = 0$, if her type is good or mixed, and propose b at $t = 0$ if her type is bad. So, when b is proposed, the principal *knows* that the agent is of bad type, and that b is the *only* feasible project. Yet he accepts b with a probability strictly lower than one. This would obviously not be possible if he lacked commitment power.

So, using probability of acceptance to screen different types of the agent is not possible when the principal lacks commitment power. Our main result however, establishes that there is always an equilibrium of the game where the principal attains his commitment payoff in the frequent-offer limit. In particular, there is a "separating" equilibrium where **delay** emerges as a costly signaling device for the agent and allows for the separation of the mixed and the bad type. This equilibrium attains the principal's optimal commitment payoff when $\lambda > \lambda^*$. We now state our main result.

Theorem 1. *There is always an equilibrium of the game in which the principal's payoff approximates his commitment payoff in the frequent-offer limit, as $\delta \to 1$. Let $t^*(\delta) := \min\{t : \alpha_g \geqslant \delta^t \alpha_b\}$. On-path behavior in the equilibria that attain the commitment payoff is as follows.*

a) *(Pooling) When $\lambda \leqslant \lambda^*$, the "pooling" equilibrium attains the principal's commitment payoff: On-path, each type of the agent proposes her favorite available project at $t = 0$. The principal accepts both g and b at $t = 0$.*

b) *(Separating) When $\lambda > \lambda^*$, the "separating" equilibrium approximates the principal's commitment payoff as $\delta \to 1$. On path:*

   * *The agent's good and mixed types propose g at $t = 0$ and the bad type stays silent until $t^*(\delta)$, at which point she proposes b.*

   * *The principal accepts g at $t = 0$ and b with probability one at $t^*(\delta)$.*

The details of the strategies and beliefs that constitute the pooling and the separating equilibria are in the Appendix. Here, we focus on the more interesting case; that of the separating equilibrium. We first argue that the on-path behavior we described indeed attains the principal's payoff from the separating mechanism. On-path, the mixed and the good types both propose g at $t = 0$, and it is accepted. This replicates the outcome of the principal's optimal commitment strategy, where g is *also* implemented (proposed and accepted) with probability one at $t = 0$, for the mixed or good types. In the optimal commitment strategy, the bad project is implemented with interior probability $\frac{\alpha_g}{\alpha_b}$ at $t = 0$ (and with probability zero thereafter). In the separating equilibrium b is implemented with probability one, but with a delay, at $t^*(\delta)$. By definition of $t^*(\delta)$, as $\delta \to 1$, we have that $t^*(\delta) \to \frac{\alpha_g}{\alpha_b}$. Thus, as $\delta \to 1$, the principal's payoff from the optimal strategy is attained by the separating equilibrium.

We now provide some intuition behind the key forces that hold the separating equilibrium together. For that, we first (informally) describe the principal's strategy (not just on-path behavior) in somewhat more detail. The principal's strategy involves:

- Accepting g whenever proposed, at any history.

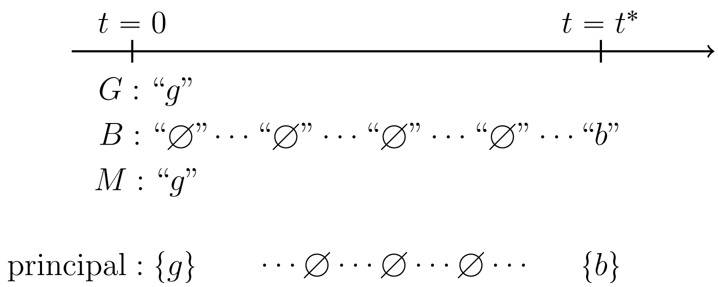

Figure 2: The timing of the proposals and accepted projects on the path of separating equilibrium of the dynamic elicitation game.

- Rejecting b if proposed at $t < t^*(\delta)$.
- If b is proposed at h_t with $t \geqslant t^*(\delta)$, accept if and only if along h_t, there is **no** $t' < t^*(\delta)$, such that b was proposed at t'.

The principal's strategy implies that if b is proposed before $t^*(\delta)$, then the principal will *never* accept b, at *any* future period. This essentially means that a type that has b *cannot* get the principal to accept it before $t^*(\delta)$. Rather, if such a type proposes b before $t^*(\delta)$, then the principal will never accept b again. So, the mixed type faces a choice: it can get g accepted right away, at $t = 0$, or wait till $t^*(\delta)$ to get b implemented. By definition of $t^*(\delta)$, it (weakly) prefers to propose g at $t = 0$. The bad type has no option but to wait by staying silent till $t^*(\delta)$. At this point she proposes her only project, and it is accepted.

Thus, delay emerges as a costly signaling device in equilibrium; it is used by the bad type to signal that she indeed **only** has the bad project. But why does the principal reject b at *any* history that involves b being proposed before $t^*(\delta)$? This is because of the principal's off-path beliefs; following such a proposal, the principal believes that it is the mixed type with probability one, and that she will propose the good project with probability one in the next period. Thus, the agent gets punished by the extremal off-path beliefs of the principal, if she ever proposes the bad project before $t^*(\delta)$.

However, ex-ante, it is not clear why this *punishment through beliefs* should be possible. Even if the principal attaches probability one to the mixed type, why does he find it optimal to always reject b given this belief? The agent still has control of the proposals after all, and the principal cannot implement something she doesn't propose. In this case, even if the principal *knows* the agent has both projects, its not obvious that he can make the agent propose g. The agent can just keep proposing b, forcing the principal to capitulate and accept it.

The intuition here is that in the complete information counterpart of our game, where it is common knowledge that the agent is of mixed type, there exists an equilibrium where, on path, the agent proposes g at $t = 0$. Consider the following strategies of the principal and the agent: the principal, irrespective of history, rejects b, and accepts g. The agent, irrespective of history, proposes g. In particular, at any history h_t, if the agent proposes g and it is rejected, then at this off-path history $h_{t+1} = (h_t, b)$, the agent's strategy is to propose g. It is easy to see that no party has a profitable one-shot deviation. For the principal, at any time period, if b is proposed, by rejecting it, he expects g to be proposed in the next period, which he would then accept. So, if he is sufficiently patient, it is optimal to reject b. For the agent, at any time period, if she proposes b, it would be rejected, and she would propose g in the next period, which would be accepted. So, her payoff from this deviation is $\delta\alpha_g$. If she doesn't deviate and proposes g in the current period, it is accepted and she gets α_g. Thus, the strategies constitute a Subgame Perfect Equilibrium.[6]

It is *this* equilibrium that our analysis leverages. So, if b is proposed at $t < t^*(\delta)$, or at any history h_t, where b was proposed and rejected at $t^{'} < t^*(\delta)$ along this history, the principal finds is optimal to reject this proposal. This is because at any such history, the principal attaches probability one to the mixed type, and expects the *complete information* equilibrium that we just described to be played. In particular, she expects the agent to propose g in the next period. This makes rejection of b sequentially rational for the principal at *any* such history, and holds this equilibrium together.[7]

We have therefore argued that off-path beliefs can be used to exploit a complete information equilibrium, and separate the mixed type from the bad. We now discuss some important features of these off-path beliefs in more detail. We also introduce a refinement in the spirit of the forward induction equilibrium from Cho (1987), that

[6] The *discreteness* of the offer space is important here. Consider the setting from Rubinstein (1982), but with one party making all the offers. Then, in the unique equilibrium, this party captures the entire surplus. However, as Van Damme, Selten, & Winter (1990) shows, any split can be supported if the offer space is discrete. A similar reasoning is at play here.

[7] There is of course another complete information equilibrium where the agent follows the *stubborn* strategy of proposing b at every history, and the principal's strategy is to accept any project at any history. However, this is not the equilibrium we exploit in our construction of the separating equilibrium.

we adapt from Admati & Perry (1987) and argue that the off-path beliefs in the separating equilibrium are consistent with this refinement.

3.2.1 Discussion of Off-Path Beliefs

In this section, we discuss a particular feature of the off-path beliefs, and its role in sustaining the *punishment play* in our separating equilibrium.

The off-path beliefs in our separating equilibrium have the feature that they *can* change even after becoming *degenerate*. To see this, recall that in the separating equilibrium, on path, if the agent is silent at $t = 0$, then the principal attaches probability one to the agent being of bad type. However, if, at $t = 1$, the agent takes the off-path action of proposing b, instead of staying silent, then the principal's belief *flips*, and he now believes, with probability one, that the agent is of mixed type.

There is nothing in the definition of Perfect Bayesian Equilibrium that rules out such *flipping*, following an off-path action. It might seem strange that beliefs are allowed to switch back from being degenerate. However, this makes more sense if we think of belief assignment as *tentative*. In a dynamic game, the fact that a deviation has happened might not reveal itself right away, because there could be deviations, that for a certain number of periods, *coincide* with the on-path behavior of *some* type. For example, in the separating equilibrium, a deviation where the agent is silent at $t = 0$ and proposes b at $t = 1$ coincides with the on-path behavior of bad type till the b is proposed. So, it seems reasonable that if the principal sees such a deviation, he is allowed to change his *tentative* belief assignment, even if it was degenerate. In light of this discussion, imposing the restriction that a degenerate belief assignment is not allowed to change seems rather restrictive. It also implies that we have limited means to punish such deviations (which mimic the on-path behavior of some type till a certain point in time), which in turn restricts the set of equilibria that we can have. In particular, we argue now, such a restriction would kill our separating equilibrium.

Intuitively, **not** having this restriction on off-path beliefs, and allowing for *flipping*, allows signaling to *continue* even after separation has occurred, and beliefs have become degenerate. In our equilibrium, separation occurs in the very first period, because the bad type takes an action (staying silent) that fully reveals her type. However, the bad type must continue waiting till $t^*(\delta)$, otherwise she is punished by unfavorable off-path beliefs. But why should signaling continue even after separation has occurred? To understand this, suppose, following separation at $t = 0$, the game is analyzed as a complete information game, where it is common knowledge that the agent is of bad type. Then, it must be that the principal always accepts b, if it is proposed following such a history. So the bad type won't stay silent till $t^*(\delta)$, she will find it optimal to propose b at $t = 1$ itself. But then, the *wait* to get b implemented is not long enough to deter the mixed type; she will find it optimal to mimic the bad type. Therefore no separation can occur.

To sum up, putting the the restriction that degenerate beliefs are **not** allowed to change, implies requiring that the continuation game is analyzed as a complete information game once beliefs are degenerate. This, as we argued above, would break the separating equilibrium. Separation occurs in this equilibrium at $t = 0$ *because* the bad type must **continue** signaling till $t^*(\delta)$. Putting this restriction seems inconsistent with allowing the agent to signal his type through *waiting*, or delay, where the principal expects one type of the agent to wait, and this is sustained by unfavorable off-path beliefs if waiting stops too early.

The literature on dynamic signaling has several examples where off-path degenerate beliefs are allowed to change, and signaling continues after beliefs have become degenerate. Examples include Noldeke & van Damme (1990), Crampton (1992), Admati & Perry (1987), and Kaya (2009). In these papers, *waiting* is less costly for the strong type, so the strong type needs to wait for a given period of time. As soon as the waiting starts, the principal attaches probability one to the strong type, if the waiting stops too early, the beliefs would change back from being degenerate. Hence, signaling for the necessary amount of time (long enough for separation to occur), is enforced by unfavorable off-path beliefs, just like in our model. However, the models in these papers, and the exact mechanism through which separation is sustained is different in our model. Closet to our paper, in terms of the use of delay as a signaling device, is Admati & Perry (1987), who analyze bargaining game where the time taken to respond to offers is a strategic variable, unlike our model. Also, their game is one of alternating offers, and as we will argue in Section 3.3, in our game, the principal's inability to make offers is important.

3.2.2 Refinements

In our game, the challenge with using refinements like Intuitive Criterion, which are defined for static signaling games, is that in our dynamic game, when a deviation occurs, the full path that the player has deviated to is not immediately observed. So, we don't really observe the full deviation, only that it has occurred. So, how do we determine the set of types that cannot possibly benefit from this deviation? Moreover, following any deviation, the principal's best response depends not just on his belief about the agent's type, but also on how various types in the support of his belief will play in the continuation game.

Finally, when an off-path action is taken at a history, it need not be the *first* off-path action along that history: It is possible that the agent took an off-path action earlier, and then *again* deviated from what the equilibrium strategy prescribes following the first deviation. For example, in our separating equilibrium, suppose the agent stays silent till $t' < t^*(\delta)$, and then proposes b at t', and then, at $t' + 1$, again proposes b. This path involves two off-path proposals: The first is obviously when b is proposed at t'. Following this deviant proposal, the equilibrium strategy for the bad type prescribes staying silent in each subsequent period, and the equilibrium strategy for the mixed type prescribes proposing g at $t' + 1$. So, the proposal of b at history $t' + 1$ is also off-path. The possibility of a deviation *within* a deviation makes the task of assigning off-path beliefs quite complex.

To deal with all these issues, we use a refinement that is in the spirit of the forward induction equilibrium concept in Cho (1987), similar to the restriction on off-path beliefs that's used in Admati & Perry (1987). This refinement *only* restricts beliefs following an off-path proposal at a history, if it is the **first** off-path proposal along that history, i.e. all previous proposals were on-path. The main idea behind this refinement is that we ask, for each type of agent, what is the maximum expected payoff that this type can get from a particular deviation. If this maximum payoff is strictly lower than her equilibrium payoff, then this deviation is **bad** for this type. Our restriction says that if a deviation is **bad** for a type S, *and* there exists a type S' for which this deviation is **not bad**, then the off-path beliefs following this deviation should assign probability zero to type S.

We first restrict attention to a particular kind of deviation here, which is particularly relevant for our separating equilibrium. Suppose the agent is silent till period $t' - 1$, and at period t', at history $h_{t'}$, she proposes b, which is an off-path proposal at this history. Moreover, it is the **first** off-path proposal along this history in the sense that along this history, silence was on-path at every $t' < t$. Of course, in general, the agent need not be silent before making her first off-path proposal. Her on-path behavior before her first off-path proposal can involve something other than staying silent. We deal with this more general case in the Appendix.

To describe the deviation we are considering formally, fix equilibrium σ^*, where σ^*_P denotes the equilibrium strategy of the principal, and σ^*_S represents the strategy of the agent of type S. Recall that $\sigma^*_S(i|h_t)$ denotes the probability with which the type S agent proposes i at h_t. Suppose that the agent is silent at every $t \leqslant t'-1$, and proposes b at t'. Let $h_{t'}$ denote the period t' history that involves the agent being silent at all $t \leqslant t'-1$. Suppose that $\{\sigma^*_S\}$ is such that along $h_{t'}$, at each $t < t'$, silence is on-path, i.e. at each $t < t'$, $\sum_{\{S|\mu_S(h_t(h_{t'}))>0\}} \sigma^*_S(\varnothing|h_t(h_{t'})) > 0$, where $h_t(h_{t'})$ is the history obtained by truncating $h_{t'}$ at t. Suppose also that at $h_{t'}$, a proposal of b is off-path, i.e. $\sum_{\{S|\mu_S(h_{t'})>0,\, b\in S\}} \sigma^*_S(b|h_{t'}) = 0$. Then, b is the first off-path proposal along history $(h_{t'}, b)$.

What is the agent's *maximum* payoff from this deviation, and how does it compare to her payoff from following her equilibrium strategy? For computing the agent's maximum payoff following her deviant proposal of b at $h_{t'}$, we consider the highest payoff that she can get in *any* PBE of the continuation game following the deviation, for any beliefs of the principal following the deviation. Formally, for any type S of the agent, we denote her *maximum* payoff from the deviation by $\alpha^*_S(h_{t'}, b)$, which we define as follows: let $\{\alpha_S(\mu, h_{t'}, b)\}_{\mu\in\Delta\{S|b\in S\}}$ be the set of equilibrium payoffs for type S in the continuation game following history $(h_{t'}, b)$. Then, $\alpha^*_S(h_{t'}, b) := \sup \bigcup_\mu \{\alpha_S(\mu, h_{t'}, b)\}$ is the highest such continuation payoff, across **all** possible beliefs.[8]

Definition 1. *Consider a deviation where the agent is silent till period $t'-1$, and proposes b at t', where b is the* ***first*** *off-path action along this history. Let $\alpha^*_S(\sigma^*_S)$ denote the payoff of type S from following her equilibrium strategy. The deviation b at history $h_{t'}$ is* ***bad*** *for type S if $\alpha^*_S(\sigma^*_S) > \delta^{t'}\alpha^*_S(h_{t'}, b)$.*

Note that the maximum payoff from the deviation we consider is simply the discounted maximum payoff following the deviation, because the agent is silent till t', so the payoff till t' is zero. We now state our restriction on off-path beliefs:

Assumption 1. *If, at a history $h_{t'}$, an off-path proposal b is observed, and the following is true: i) b is the first off-path proposal along the history $(h_{t'}, b)$, and ii) there is a type S for which this deviation is* ***bad****, and a type S', for which this deviation is* ***not bad****. Then, following this deviation, the beliefs attach probability zero to type S.*

[8]With the restriction that beliefs can only assign positive probability to types that contain b.

Note that the above assumption imposes no restrictions when the deviant offer of b is either **bad** for all possible types that can propose b, or **not bad** for all possible types that can propose b. We now argue that beliefs in our separating equilibrium are consistent with this restriction. But before that, we prove a useful Claim.

Claim 1. *Following an off-path proposal of b at any history $h_{t'}$, the maximum payoff from this deviation for any type S such that $b \in S$, is $\alpha^*_S(h_{t'}, b) = \alpha_b$. So, the maximum payoff from this deviation across all continuation equilibria is the payoff from this deviant proposal of b being accepted with probability one.*

Proof. Recall the *pooling equilibrium*, where the strategy of each type of the agent is to propose her favorite *available* project at *any* history, and the principal's strategy is to accept with probability one, at *every* history, any project that's proposed. This equilibrium exists for any belief of the principal. So, following *any* deviation, and any belief of the principal following that deviation, the pooling equilibrium is an equilibrium of the continuation game. Now, fix an equilibrium and consider a history $h_{t'}$ where b is an off-path proposal. By deviating and proposing b at $h_{t'}$, obviously no type of the agent can do better than if b is accepted. And if the pooling equilibrium is played following this deviation, then b is accepted with probability one. So, $\alpha^*_S(h_{t'}, b) = \alpha_b$. □

Claim 2. *The beliefs in our separating equilibrium satisfy Assumption 1.*

Proof. The only deviation we care about is if b is proposed *before* $t^*(\delta)$. We know from 1 that the agent's maximum payoff from such a deviation is $\delta^{t'}\alpha_b$. Recall that $t^*(\delta) = \min\{t|\alpha_g \geqslant \delta^t\alpha_b\}$. So, since the equilibrium payoff of the mixed type is α_g, the equilibrium payoff of the bad type is $\delta^{t^*(\delta)}\alpha_b$, and $t' < t^*(\delta)$, for **both** the mixed and bad types, this maximum deviation payoff of $\delta^{t'}\alpha_b$ is *strictly* higher than their equilibrium payoffs. So, this deviation is **not bad** for either of these types, and therefore our off-path belief that following such a deviation, the principal attaches probability one to the mixed type, satisfies Assumption 1. □

We now give the example of *another* separating equilibrium, with a longer waiting time, that does not survive this restriction on beliefs.

Claim 3. *A separating equilibrium with a wait time longer than $t^*(\delta) + 1$ does not satisfy Assumption 1.*

Proof. Consider a separating equilibrium where instead of $t^*(\delta)$, the bad type is expected to stay silent till $t^{**} > t^*(\delta) + 1$ to signal her type. Like our original separating equilibrium, if b is proposed before t^{**}, the principal rejects it because his off-path beliefs are that its the mixed type and will propose g in the next period. These off-path beliefs however, do not satisfy Assumption 1. To see this, suppose b is proposed at t' such that $t^*(\delta) < t' < t^{**}$. Such at t' exists because $t^{**} > t^*(\delta) + 1$. We know that the maximum payoff from such a deviation is $\delta^{t'}\alpha_b$, which is the payoff from this proposal being accepted. From the definition of $t^*(\delta)$, this is strictly lower than α_g, the equilibrium payoff of the mixed type, but is strictly higher than $\delta^{t^{**}}\alpha_b$, the equilibrium payoff of the bad type. So, this deviation is **bad** for the mixed type, but **not bad** for the bad type. Thus, following such a deviation, the principal should attach probability one to the bad type as per Assumption 1. Our off-path beliefs therefore do not satisfy the assumption. □

3.3 Comparison with Dynamic Delegation Game

As we discussed in the previous section, in our separating equilibrium, signaling continues for multiple periods, *after* beliefs have become degenerate. For this to happen, it seems crucial that at **no** point, while signaling is going on, can the principal *intervene*, and **allow** the bad project, i.e. *commit* to accepting b if it is proposed, effectively delegating the decision to the agent. This seems important because if the principal had this ability, then right after observing silence at $t = 0$, he would *allow* b. This in turn would mean that the mixed type cannot be prevented from mimicking the bad type.

Motivated by this, and to further understand the role of the *lack of proposal power* for the principal, where he can only respond to the agent's proposals, but cannot propose, or *allow* a project himself, we compare our extensive form to one where the principal does have proposal power. We see that here, the principal's ability to propose prevents delay from emerging as a signaling device, and no separation is possible in equilibrium.

The *dynamic delegation game*, is one where in each period, the principal chooses a permission set $A_t \subseteq N$. The agent can choose a feasible project from A_t, or stay silent and choose nothing. If the agent chooses nothing, the game moves to the next period, where once again, the principal chooses a permission set. The principal cannot commit to what permission sets he will choose in future. This game is studied by Li (2024), who shows that the ability to *revise* the permission set in each period proves to be disastrous. His main result is that **the unique equilibrium outcome of this game is that the principal gives full discretion to the agent at the outset, i.e at $t = 0$, the principal allows all possible projects.**

So, there is **no** equilibrium where any signaling happens and the principal is able to learn the agent's type over time. For the formal proof, refer to Li (2024), but we will provide some intuition here. The culprit here is the principal's ability to *allow* the bad project, if he is (sufficiently) convinced that the good project is not feasible. Suppose we attempt to replicate the outcome of our separating equilibrium in an equilibrium of this game. There, the bad type uses delay as a signaling device. In this principal-proposes game, can the principal use delay as a *screening* device?

Consider a strategy of the principal where he chooses the permission set $A_t = \{g\}$ at all $t < t^*(\delta)$, and at $t^*(\delta)$, he chooses $A_{t^*(\delta)} = \{g, b\}$. So the mixed type indeed finds it optimal to accept g at $t = 0$, and the bad type waits till $t^*(\delta)$, when b is permitted. It is the *principal's* sequential rationality that's the problem. Given the agent's best response, this strategy of the principal is not sequentially rational; at $t = 0$, if g is not accepted, the principal believes that its the bad type with probability one, and therefore would permit b in the very next period. His sequential rationality prevents him from waiting till $t^*(\delta)$ before allowing the bad project. Observe that if the principal could **commit** to his strategy, then he can use such a strategy to screen through delay. However, without commitment, his sequential rationality gets in the way.

In fact, no separation along the equilibrium path is possible. The idea is that if the principal does not allow both projects with probability one, at $t = 0$, then, he cannot allow b with probability one at any $t' > 0$, because if he does, and is δ is sufficiently high, both the mixed and bad types would stay silent at $t' - 1$, and choose b when it is permitted at t'. But then, the principal is better off allowing b at $t' - 1$ itself. So, if full discretion is not given at $t = 0$, then, in each subsequent period, the game must continue to the next with positive probability. Therefore, there exists an equilibrium path along with there is arbitrarily long delay in reaching an agreement. Along this path, the principal must eventually realize that g is either not feasible, or the agent will not propose it. He is therefore better off simply allowing b at some point along this path, which is a contradiction.

To sum up, in this dynamic delegation game, the principal has the ability to commit to a set of projects that he will approve in any period. This short-term commitment has adverse consequences for the principal in the dynamic game, as it puts the burden of sequential rationality on him. In our agent-proposes game, this burden lies with the agent: *she* must not propose b before $t^*(\delta)$ to avoid being *punished* by unfavorable off-path beliefs. In the dynamic delegation game, the principal can, at any point, simply *permit* b; he does not have to wait for the agent's proposal. So punishment though off-path beliefs is not possible here, as the principal cannot form off-path beliefs in response to his own actions. Finally, note that since permitting both projects at $t = 0$ is the unique equilibrium outcome here, the principal is worse off than even the static delegation game, where he can at least commit upfront to **not** permitting b if it is highly likely that the agent is of mixed type.

3.3.1 Comparison With Other Extensive Forms

We mentioned in the introduction that our dynamic elicitation game does better that **static constrained delegation**, even though in a dynamic interaction, the principal faces a loss of commitment power with respect to the static delegation game, where he can commit to what projects he would approve upfront. We show this formally in Section A.3 of the Appendix.

Coming back to dynamic interactions, there could be other, more complex forms of interaction between the principal and the agent. The principal and the agent could make alternating offers for instance. Or, there could be some rule that determines who gets to make an offer at what history. We leave a more formal analysis of this what can be achieved in equilibrium in these more complex extensive forms for future work, but we can make one preliminary point: ***no** reasonable extensive form in this setting can have an equilibrium that gives the principal a strictly higher payoff than the principal-optimal equilibrium in our dynamic elicitation game.*

Recall that we are operating in a setting where the agent's private information is about which projects are *feasible*, and **only** feasible projects can be implemented. So any reasonable extensive form in this setting must have two features. Firstly, if it is the principal's turn to make an offer, then the agent can only *accept* offers involving projects that are feasible. For example, if the principal's proposal takes the form of a permission set, then the agent can only choose a project that's feasible from this set. The second feature is that if its the agent's turn to propose, she can only propose feasible projects.

We formally define these extensive forms Section A.4 of the Appendix and argue that none of these other extensive forms can do better than ours, from the principal's point of view. The main idea is that by committing to a strategy in any of these extensive forms, the principal cannot do strictly better than he can by committing to a strategy in the dynamic elicitation game. And because, as $\delta \to 1$, the commitment payoff in our game can be achieved in equilibrium, it follows that no equilibrium of any of these extensive form can be, for the principal, strictly better than the principal-optimal equilibrium of our game.

4 General Case: N Projects

Suppose now, instead of two, we have $N \geqslant 2$ possible projects, so the set of all *possible* projects is now $\mathcal{N} \equiv \{1, 2, \ldots, N\}$ where project i corresponds to (α_i, π_i). As before, only the agent knows which projects are *available*; her *type* is $S \subseteq \mathcal{N}$ representing the set of available projects. The agent's type is drawn from $\mathcal{S} \equiv 2^{\mathcal{N}}$ according to the probability distribution $\mu : \mathcal{S} \to [0, 1]$. We make he following assumption about the conflict of interest between the principal and the agent.

Assumption 2. *(Conflicting preferences) The set of projects $\mathcal{N}$ satisfies*

$$\pi_1 > \pi_2 > \ldots > \pi_{N-1} > \pi_N > 0;$$

$$\alpha_N > \alpha_{N-1} > \ldots > \alpha_2 > \alpha_1 > 0.$$

The above assumption says that the preferences of the principal and the agent are diametrically opposed, with the principal preferring lower indexed projects and the agent preferring higher indexed projects. Of course, this is not without loss; there are other possibilities. The principal and the agent could, for instance, have the same ranking over a subset of projects. We assume however, that their rankings over the set of possible projects are exactly opposite.

With more than two projects, it is difficult to characterize the principal-optimal equilibrium. We cannot say much here, but certain insights from the two-project case continue to hold. We still have separating equilibria where projects that are more preferred by the principal are proposed, and implemented earlier, and projects that the principal dislikes, are proposed with a delay, by types of the agent who **only** have these projects. So, the separating equilibria here are similar to the separating equilibrium with two projects, except that different degrees of separation are now possible. For example, if $N = 3$, then there could be a partially separating, partially pooling equilibrium, where projects 1 and 2 are accepted whenever proposed, and project 3 is only accepted if proposed after a prescribed delay. If it is proposed earlier, the principal believes that the agent is of a type that has 1, his favorite project, and will propose it in the next period. We could also have perfectly separating equilibria, where project 1 is accepted whenever proposed, project 2 only if it is proposed with a delay of t_2^*, and project 3 only if it proposed with a delay of t_3^*, where $t_3^* > t_2^*$.

So, informally, a **separating equilibrium** in this setting specifies, for each project i, a prescribed delay t_i^*, where $t_{i'}^* < t_i^*$ for $i' < i$, i.e. the delay for lower indexed projects, or ones that are more preferred by the principal, is lower. If a project i is proposed before t_i^*, it is rejected, because the principal believes that the agent has a project that he prefers to i. These t^*'s are such that a type of an agent that has $i' < i$ will prefer to propose i' at $t_{i'}^*$, rather than wait till t_i^* to get i implemented. Observe that a fully pooling equilibrium, where $t_i^* = 0$ for all i, also lies in this class of equilibria. The details of the separating equilibria that we will be interested in are in the Appendix.

Now, we describe such assumptions such that if the parameters satisfy these assumptions, then the principal's commitment payoff can be attained in equilibrium, *by* a separating equilibrium.

Assumption 3. *(Linear payoffs) Any two projects $i, j \neq 1$ satisfy*

$$\frac{\pi_1 - \pi_i}{\alpha_i - \alpha_1} = \frac{\pi_1 - \pi_j}{\alpha_j - \alpha_1}.$$

The above assumption says that the payoffs for the projects lie on a line. This assumption gives some structure to the problem that makes solving for the optimal commitment strategy easier.

Assumption 4. *(Nested types) The probability distribution over the agent's types μ is such that for any $S, S' \in \mathcal{S}$ with $\mu(S), \mu(S') > 0$, either $S \subseteq S'$ or $S' \subseteq S$.*

Assumption 3 requires the set of possible types to be nested in a way that a type of the agent is either a subset or a superset of any other type. This assumption provides

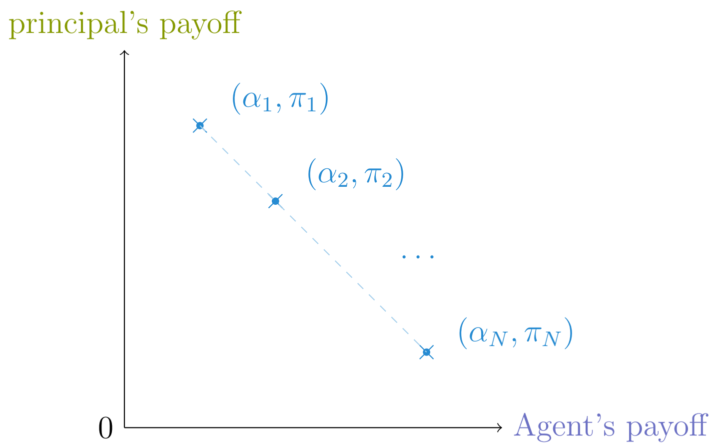

Figure 3: The project space with $N > 2$ projects under Assumptions 2 and 3. We see the conflicting preferences of the principal and agent, and the linear payoffs of the projects.

a structure to possible types and simplifies the incentives. Under Assumption 3, there can be at most one type with n projects for each $n \in \{1, 2, \ldots, N\}$.

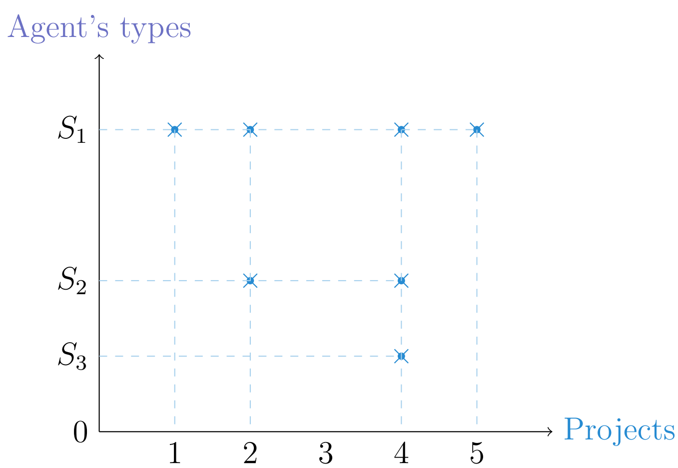

Figure 4: The type space $\mathcal{S}$ with $N > 2$ projects under Assumption 4 where project i refers to (α_i, π_i). The type space is nested such that if we take any two types, one would be a subset of the other.

When the parameters $\mathcal{N}$ and μ satisfy Assumptions 2, 3, and 4, we refer to this restricted type space with the restricted payoff structure as *nested linear type space*. The nested linear type space reduces the number of incentive compatibility constraints to at most $(N - 1)$, simplifying the problem significantly.

Under these regularity conditions provided by Assumptions 2, 3, and 4, our main result extends to the general model, and the commitment payoff is always attainable in an equilibrium of the dynamic elicitation game.

Theorem 2. *In the nested linear type space, there exists an equilibrium of the dynamic elicitation game that attains the principal's commitment payoff as* $\delta \to 1$.

The main idea behind the proof is that we can divide solving for the principal-optimal mechanism into two parts. We first establish that in any optimal mechanism, each type's each possible report generates the same expected payoff v for the agent. Then, we can solve the optimization problem for a fixed value of v for each type. Combined with the fact that the differences between payoffs are linear, the optimal mechanism takes a very clean separating structure. The optimal mechanism can then be replicated in equilibrium with similar strategies as in the separating equilibrium in the two project case.

We end this section with an example that shows that our assumption of a nested linear type space is sufficient but not necessary: there are examples of type spaces outside this class where the commitment payoff can be achieved in equilibrium. There are three possible projects, $\mathcal{N} = \{1, 2, 3\}$, and three equally likely types in the support of μ with $\mathcal{S} = \{\{1, 2\}, \{2\}, \{2, 3\}\}$. The payoffs are:

$$\pi_1 = 8, \pi_2 = 3, \pi_3 = 1$$

$$\alpha_1 = 3, \alpha_2 = 8, \alpha_3 = 9$$

Note that we are outside the linear nested type space as the types are not nested and the payoffs are not linear. This type space can be thought of as augmenting our two-project case with a type where the bad project is paired with an even worse project. It can be shown that the optimal mechanism is as follows:

- From type $\{1, 2\}$, project 1 is implemented with probability one
- From type $\{2\}$, project 2 is implemented with probability $\frac{3}{8}$.
- From type $\{2, 3\}$, project 2 is implemented with probability one.

The structure of an equilibrium that attains the payoff from this mechanism is similar to the separating equilibrium but it exhibits a novel signaling opportunity. The principal always accepts project 1 and never accepts project 3. If project 3 is proposed at $t = 0$, then project 2 is accepted with certainty at $t = 1$. Otherwise, project 2 is only accepted with a delay at $t^* = \min\{t | \delta^t \leqslant \frac{3}{8}\}$. We should highlight that even though project 3 is never implemented, its proposal acts as a screening device and the agent has an opportunity to signal her type by proposing redundant projects.

5 Conclusion

In this paper, we studied a dynamic principal-agent problem where the agent is privately informed about the feasibility of projects, and the interests of the parties are not aligned. The agent makes proposals over time, and the principal has the authority to approve or reject these proposals, but cannot commit to his responses to *future* proposals.

We ask how well the principal can do here. He lacks commitment power and *proposal power*: he can only accept projects that the agent proposes, which results in a reduced level of control over what is implemented. We might expect that that the agent can easily *hide* principal-preferred projects by never proposing them. Anticipating that his preferred projects may never be proposed, the principal would in turn capitulate and accept the agent-preferred projects when they are proposed. We show however, that with two projects, there is always an equilibrium of our dynamic game where the agent proposes the principal-preferred project in the very first period if its available, and only proposes the agent-preferred project if the principal-preferred project is not available, after *waiting* for a prescribed amount of time. In fact, this equilibrium attains the principal's commitment payoff. We argue that it is in fact the inability to make proposals that enables the principal to *wait*, and for costly delay to emerge as a signaling device in equilibrium. For more than two projects, we identify sufficient conditions on parameters under which the commitment result still holds.

We also compare our dynamic elicitation game with another extensive form studied in the literature, and our main observation from this comparison is that not every kind of dynamic interaction allows the principal to learn about the agent's type over time. In the **dynamic delegation game** that we compare our game to, the principal's ability to to *allow* projects, or to delegate the choice to the agent, prevents signaling through delay, and the principal capitulates and provides full discretion to the agent in the unique equilibrium outcome.

Finally, our results and the above comparison have implications for organizational design, as discussed in the Introduction. Our main implications are for a setting where when the manager privately knows which decisions are feasible, the decisions that are feasible cannot be implemented without the manager's cooperation.[9] This could be because assessing both the feasibility of the decision, as well as implementing it, requires specialized knowledge that only the manager has. However, the manager is motivated by empire building: She wants to implement *flashy* projects that increase her own influence within the organization, rather than projects that are in the best interest or shareholders. We find that in such a setting, adopting a bottom-up approach, where the CEO elicits proposals from the manager, may be better at curbing these empire building plans than issuing top-down commands, restricting what the manager can do.

[9] Unless she proposes or chooses it.

References

Admati, A. R., & Perry, M. (1987). Strategic delay in bargaining. *The Review of Economic Studies*, *54*(3), 345–364.

Alsonso, R., & Matouschek, N. (2008). Optimal delegation. *The Review of Economic Studies*, *75*(1), 259–298.

Armstrong, M., & Vickers, J. (2010). A model of delegated project choice. *Econometrica*, *78*(1), 213–244.

Cho, I. K. (1987). A refinement of sequential equilibrium. *Econometrica*, *55*(6), 1367–1389.

Crampton, P. C. (1992). Strategic delay in bargaining with two-sided uncertainty. *The Review of Economic Studies*, *59*(1), 205–225.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies*, *69*(4), 811–838.

Holmström, B. (1984). On the theory of delegation. *In Bayesian Models in Economic Theory,eds. M. Boyer and R. Kihlstrom*, 115–141.

Kaya, A. (2009). Repeated signaling games. *Games and Economic Behavior*, *66*(2), 841–854.

Li, W. (2024). Full discretion is inevitable. *Working Paper*.

Melumad, N. D., & Shibano, T. (1991). Communication in settings with no transfers. *The RAND Journal of Economics*, *22*(2), 173–198.

Nocke, V., & Whinston, M. D. (2010). Dynamic merger review. *Journal of Political Economy*, *118*(6), 1200–1251.

Nocke, V., & Whinston, M. D. (2013). Merger policy with merger choice. *American Economic Review*, *103*(2), 1006–33.

Noldeke, G., & van Damme, E. (1990). Signalling in a dynamic labour market. *The Review of Economic Studies*, *57*(1), 1–23.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, *50*, 97–109.

Schneider, J. (2015). Persuasion, pandering, and sequential proposal. *Working Paper*.

Van Damme, E., Selten, R., & Winter, E. (1990). Alternating bid bargaining with a smallest money unit. *Games and Economic Behavior*, *2*(2), 188–201.

Appendices

A Proofs

A.1 Proof of Proposition 1

To show that in either case ($\lambda > \lambda^*$ or $\lambda \leqslant \lambda^*$), the commitment strategy that we described is indeed optimal, we proceed in **two** steps.

1. We first define a class of static, stochastic mechanisms, and show that ability to commit to a mechanism in this class would get the principal at least as high a payoff as he can get by committing to a strategy in the dynamic game. So, this class of mechanisms acts as an *upper bound* for what the principal can achieve by committing to strategy in the dynamic game.

2. Then, we compute the principal optimal mechanism in this class, and show that the outcome of this, and therefore the principal's payoff in each case ($\lambda > \lambda^*$ and $\lambda \leqslant \lambda^*$), corresponds to that of commitment strategies we described. This would complete the proof that the strategies we described are indeed optimal.

We now define a class of static, stochastic mechanisms with type-dependent message spaces. In a mechanism in this class, the message space is type dependent, and the set of messages that a type S of the agent can send is $M(S) = 2^S$. So, the agent can only report *subsets* of her type, or the set of *available* projects. [10] A mechanism is a tuple (M, q), where $M = \bigcup_{S \in \mathcal{S}} M(S)$ is the set of all possible messages, and $q : \mathcal{S} \to \Delta(S \cup \varnothing)$ is the outcome function. When the agent makes a report, the principal maps that report to a probability of implementation of *each* project in the report. herefore, when S is reported, only projects *in* S can be implemented, or no project at all, as captured by $\varnothing$. If no project is implemented, the players obtain the status quo payoff, which is zero for both the principal and the agent. For any type $S \in \mathcal{S}$ and any project $i \in S$, $q_S(i)$ represents the probability of implementing project i when the type S is reported.[11] We define a mechanism to be *incentive compatible* (IC) if no type finds it optimal to report a strict subset.

[10] This includes the empty set.

[11] So, $1 - \sum_{i \in S} q_{Si}$ is the probability of not implementing any project when S is reported.

Every mechanism determines an allocation, which is a vector $\{q_S(i)\}_{S\in\mathcal{S},i\in S}$. A *feasible* allocation is one where for each type S, we have: $q_S(i) \in [0,1]$ for each $i \in S$ and $\sum_{i\in S} q_S(i) \leqslant 1$. The principal-optimal mechanism maximizes the principal's payoff by choosing implementation probabilities for projects in each type, subject to feasibility and incentive compatibility constraints.

$$\max_{q_G(g),q_B(b),q_M(g),q_M(b)\in[0,1]} \quad \mu_G q_G(g)\pi_g + \mu_B q_B(b)\pi_b + \mu_M q_M(g)\pi_g + \mu_M q_M(b)\pi_b$$

$$\text{subject to} \qquad \textstyle\sum_{i\in S} q_S(i)\alpha_i \geqslant \sum_{i'\in S'} q_{S'}(i')\alpha_{i'} \qquad (\mathrm{IC}_{SS'})$$

$$q_M(g) + q_M(b) \leqslant 1$$

where $\mathrm{IC}_{SS'}$ denotes the IC constraint for type S to not report type $S' \subseteq S$. The second constraint is just the feasibility constraint for the implementation probabilities when the mixed type is reported.

Proposition 2. *Fix a commitment strategy of the principal, and a best response of the agent to this strategy. Then, there exists an incentive compatible mechanism in the above class, such that the principal's payoff from committing to this mechanism is at least as high as her payoff from the commitment strategy.*

Proof. Fix a commitment strategy σ_P^* of the principal and a best response $\{\sigma_S\}_{S\in\mathcal{N}}$ of the agent, where σ_S is the best response of the agent of type S. An *outcome* of the game is represented by (i,t), which means that project i is implemented at time t. No project ever being implemented is also a possible outcome. The commitment strategy and the agent's best response together induce a *for each type*, a probability distribution over outcomes.

Fix a type S. At any history h_t, $\sigma_S(i|h_t)$ is the probability with which $i \in S \cup \varnothing$ is proposed by an agent of type S, and $\sigma_P^*(i|h_t)$ is the probability with which the principal accepts this offer. We define the probability of any history arising, along the equilibrium path, inductively. We denote period 0 history by h_0, so at $t=1$, for any $h_1 = (h_0, i)$, where $i \in S \cup \{\varnothing\}$. We define $\nu(h_1) := \sigma_S(i|h_0)(1-\sigma_P^*(i|h_0))$, which is just the probability that i was proposed at $t=0$ but not accepted, and thus the probability of history h_1 at $t=1$. This is clearly a number in $[0,1]$. Given that we have defined $\nu(h_t)$ $\forall$ $h_{t'}$, $t' \leqslant t$, and any $h_{t+1} = (h_t, i)$ for some $i \in S \cup \{\varnothing\}$, we have that $\nu(h_{t+1}) = \nu(h_t)\sigma_S(i|h_t)(1 - \sigma^*_P(i|h_t))$.

We define the probability of outcome (i, t) as

$$p_S(i,t) := \sum_{h_t \in \mathcal{H}_t} \nu(h_t)\sigma_S(i|h_t)\sigma^*_P(h_t, i)$$

for all $i \in S$. It can be verified that the sum of probabilities for all outcomes in which a project *is* implemented,

$$\sum_{t=0}^{\infty}\sum_{i \in S} p_S(i,t) \leqslant 1,$$

where the probability of the outcome that *no* project is ever implemented is

$$1 - \sum_{t=0}^{\infty}\sum_{i \in S} p_S(i,t).$$

Thus, for any type S, the strategy and best response induce a probability distribution over outcomes.

We now define an *induced* conditional probability distribution, which specifies, for each type S, the (discounted) probability with which project i is proposed. For any S, probability of implementation of $i \in S'$ is

$$f^I_S(i) = \sum_{t=0}^{\infty} \delta^t p_S(i,t).$$

Now, we construct the mechanism that achieves the principal's payoff from the above commitment strategy. Consider the mechanism that maps any report S, to a probability of implementation $f^I_S(i)$ for each $i \in S$, where $1 - \sum_{i \in S} f^I_S(i)$ is the probability with which **no** project is implemented, when S is reported. Clearly, this mechanism is IC, because if, in response to the commitment strategy, no type could benefit from deviating to the best response of another type, then, given this mechanism, no type can be strictly better off by reporting a strict subset. And this mechanism clearly gives the principal the same expected payoff as the commitment strategy, because it induces the same type dependent distribution over outcomes. $\square$

We now solve for the principal-optimal mechanism and show that either the separating or the pooling equilibrium always achieves the principal's payoff from the optimal mechanism. Before we solve for the principal-optimal mechanism, we make a few simplifying observations about some properties that must be true for an optimal mechanism.

Observation 1: *In an optimal mechanism, we must have* $q^*_M(g) + q^*_M(b) = 1$. When the mixed type is reported, the probabilities of implementing the good and the bad projects must sum up to 1 in an optimal mechanism. Suppose not, i.e. $q^*_M(g) + q^*_M(b) < 1$. Then, we can increase both $q^*_M(g)$ or $q^*_M(b)$ slightly, and have new implementation probabilities $(q^{**}_M(g), q^{**}_M(b)) > (q^*_M(g), q^*_M(b))$ and $q^{**}_M(g) + q^{**}_M(b) < 1$. It must be the case that the IC constraints involving the mixed type still hold, as

$$q^{**}_M(g)\alpha_g + q^{**}_M(b)\alpha_b > q^*_M(g)\alpha_g + q^*_M(b)\alpha_b \geqslant q_G(g)\alpha_g$$

$$q^{**}_M(g)\alpha_g + q^{**}_M(b)\alpha_b > q^*_M(g)\alpha_g + q^*_M(b)\alpha_b \geqslant q_B(b)\alpha_b$$

From these new implementation probabilities, the principal obtains a strictly higher payoff. So, $q^*_M(g) + q^*_M(b) < 1$ cannot be part of an optimal mechanism.

Observation 2: *The incentive compatibility constraint for the mixed type to not report the good type,* IC_{MG}*, is redundant.* We know from the previous observation that when the mixed type is reported, the probabilities of implementing projects sum up to 1. It implies that the payoff of the mixed type, from reporting truthfully, will be at least α_g. On the other hand, her payoff from reporting the good type is at most α_g as $q_G(g) \in [0, 1]$. Thus, the mixed type is always weakly better off by reporting truthfully than by pretending to be the good type, and IC_{MG}, is redundant.

Observation 3: *In any optimal mechanism, we must have* $q_G(g) = 1$. Since IC_{MG} is redundant, and there is no type other than the nixed type that can report the good type, therefore when the good type is reported, an optimal mechanism must implement the good project with certainty.

Given the above *observations*, the problem of finding the optimal mechanism reduces to that of choosing q_{Mg} and q_{Bb} to maximize the principal's expected payoff, subject to IC_{MB}. This is because the other IC constraints are redundant, $q_{Gg} = 1$, and $q_M(g) + q_M(b) = 1$, so choosing $q_M(g)$ and $q_B(b)$ pins down the optimal mechanism.

The lone IC constraint, IC_{MB}, represents the trade off that the principal faces in implementing the good project with positive probability from the mixed type. If $q_M(g) > 0$, the principal will have to set $q_B(b) < 1$, so that the mixed type doesn't imitate the bad type. We now define two mechanisms, and it turns out that one of them is always *an* optimal mechanism.

Definition. *The pooling mechanism implements the bad project from the agent's bad and mixed types:* $q^*_G(g) = 1, q^*_B(b) = 1, q^*_M(g) = 0, q^*_M(b) = 1$.

The separating mechanism implements the good project from the mixed type and the bad project from the bad type with an interior probability: $q^*_G(g) = 1, q^*_B(b) = \frac{\alpha_g}{\alpha_b}, q^*_M(g) = 1, q^*_M(b) = 0$.

It is easy to see that both these mechanism are IC. In the *pooling* mechanism, the outcome when the type is mixed, is same as the outcome when the type is bad. In both cases, the bad project is implemented with probability one. So, mixed type is *pooled* with the bad. On the other hand, in the *separating* mechanism, the outcome when the type is mixed is different from the outcome when the type is bad. In one case, the good project is implemented with probability one, and in the other, the bad project is implemented with probability $\frac{\alpha_g}{\alpha_b}$. So, this mechanism *separates* the mixed and bad types.

Recall that $\lambda = \frac{\mu_o(M)}{\mu_0(B)}$ is the likelihood ratio of mixed type compared to bad type, and $\lambda^* = \frac{(1-\frac{\alpha_g}{\alpha_b})\pi_b}{(\pi_g-\pi_b)}$.

Proposition 3. *Either the pooling or the separating mechanism is always optimal.*

a) When $\lambda < \lambda^*$, the pooling mechanism is optimal.

b) When $\lambda > \lambda^*$, the separating mechanism is optimal

c) When $\lambda = \lambda^*$, any mechanism with $q_{Gg} = 1$, $q_{Mg}+q_{Mb} = 1$, and $q_{Mg}\alpha_g+q_{Mb}\alpha_b = q_{bB}\alpha_b$ is optimal. In particular, both the pooling and separating mechanisms are optimal.

Proof. Given that $q_{Gg} = 1$ and $q_{Mg} + q_{Mb} = 1$, our maximisation problem reduces to:

$$\max_{q_{Mg},q_{Bb}\in[0,1]} \quad \mu_G\pi_g + \mu_B q_{Bb}\pi_b + \mu_M q_{Mg}\pi_g + \mu_M(1 - q_{Mg})\pi_b$$

$$\text{subject to} \quad q_{Mg}\alpha_g + (1 - q_{Mg})\alpha_b \geqslant q_{bB}\alpha_b$$

In an optimal mechanism, we must also have that $q_M(g)\alpha_g + (1 - q_M(g))\alpha_b = q_B(b)\alpha_b$. To see this, suppose the inequality is strict. Let $q_M(g)' = q_M(g) + \varepsilon$, and $q_M(b)' = 1 - q_M(g)' = q_M(b) - \varepsilon$, where $\varepsilon > 0$. For ε small enough, IC_{MB} is still satisfied and the principal's expected payoff increases by $\varepsilon\mu_M(\pi_g - \pi_b)$. We can therefore substitute $q_M(g)\alpha_g + (1 - q_M(g))\alpha_b = q_b(B)\alpha_b$ into our objective function, and we get

$$\mu_G\pi_g + \mu_B\pi_b(1 - q_M(g)(1 - \frac{\alpha_g}{\alpha_b})) + \mu_M q_M(g)\pi_g + \mu_M(1 - q_M(g))\pi_b$$

$$= \mu_G\pi_g + \mu_M\pi_b + q_M(g)\{\mu_M(\pi_g - \pi_b) - \mu_B\pi_b(1 - \frac{\alpha_g}{\alpha_b})\}$$

The only choice variable is $q_M(g)$ now, and whether the above expression is increasing or decreasing in $q_M(g)$ depends on the sign of it's coefficient, $\{\mu_M(\pi_g - \pi_b) - \mu_B\pi_b(1 - \frac{\alpha_g}{\alpha_b})\}$. If the coefficient is strictly positive, then the optimal mechanism has $q_M(g) = 1$. Also, because IC_{MB} holds with equality, we have $q_B(b) = \frac{\alpha_g}{\alpha_b}$ in the optimal mechanism. A bit of rearranging gives us that $\{\mu_M(\pi_g - \pi_b) - \mu_B\pi_b(1 - \frac{\alpha_g}{\alpha_b})\} > 0$ is equivalent to $\lambda > \lambda^*$. Similarly, if $\lambda < \lambda^*$, the optimal mechanism has $q_M(g) = 0$, and therefore $q_B(b) = 1$. If $\lambda = \lambda^*$, principal's expected payoff is constant in $q_M(g)$ and therefore any $q_M(g) \in [0, 1]$ is optimal, with $q_B(b)$ again being determined by the equality of IC_{MB}.

□

Observe that the outcome of the separating mechanism corresponds exactly to the outcome of the principal's optimal commitment strategy when $\lambda > \lambda^*$, and the outcome of the pooling mechanism corresponds exactly to the outcome of the principal's optimal commitment strategy when $\lambda \leqslant \lambda^*$. This completes the proof of Proposition 1.

A.2 Proof of Theorem 1

We argued in the main text that the separating equilibrium attains the principal's optimal commitment payoff as $\delta to 1$, when $\lambda > \lambda^*$ and the pooling equilibrium attains the principal's optimal commitment payoff when $\lambda \leqslant \lambda^*$.

We now describe the strategies in the pooling and the separating equilibria formally, and show that they indeed constitute an equilibrium. First, we establish some notation. Recall that for any time period t, the set of all possible period t histories is denoted by $\mathcal{H}_t$, where $\mathcal{H}_t = (N \cup \varnothing)^t$. The representative period t history is denoted by $h_t \in \mathcal{H}_t$. The action (or proposal) space of an agent of type S at any history is given by $A_S(h_t) = A_S = \{S \cup \varnothing\}$, where $\varnothing$ represents remaining silent. An element of $A_S(h_t)$ is given by a_S^t. Given a history h_t, and a proposal in $(N \cup \varnothing)$ by the agent, the principal can either *accept*, or *reject* this proposal.

A behavior strategy maps histories and types into action spaces. For the agent of type S, $\sigma_S(i|h_t)$ denotes the probability of proposing project i at history h_t. For the principal, $\sigma_P(i|h_t)$ denotes the probability of accepting proposal $i \in N$ at history h_t. If at h_t, agent is silent, then $\sigma_P(\varnothing|h_t) = 0$, as there is no project to accept. At any history h_t, we denote the probability the principal attaches to type S by $\mu_S(h_t)$. If, following h_t, i is proposed, the updated beliefs are given by $\mu_S(h_t, i)$ for each S, and by $\mu_S(h_t, \varnothing)$ if the agent is silent at h_t. For any t and any $t' < t$, let $h_t(t') \in (\mathcal{N} \cup \varnothing)$ be the proposal at period t', along this history h_t. We denote by $h_{t-1}(h_t)$ the period $t-1$ history obtained by removing proposal $h_t(t-1)$ from h_t, so that $h_t = (h_{t-1}(h_t), h_t(t-1))$.

Our solution concept is Perfect Bayesian Equilibrium, as defined in Fudenberg and Tirole (1991). We want to highlight that here, something stronger than Bayes' Rule is used to update beliefs following any proposal at any history. To understand this, fix any history h_t. Even if this is a history that arises with probability zero along the equilibrium path, beliefs following a proposal i at this history are updated using Bayes' Rule if $\exists$ a type S such that $\mu_S(h_t) > 0$ and $\sigma_S(i|h_t) > 0$. Beliefs are allowed to be *completely arbitrary* only if given h_t, the proposal made by the agent had probability zero, according to the agent's strategy. However, even when following a proposal i at history h_t, beliefs are allowed to be completely arbitrary, they must still have support in $\{S \subseteq \mathcal{N} | i \in S\}$, i.e. the set of types that have i, because only these types can possibly propose i. This is in contrast to models without hard evidence.

A.2.1 Pooling equilibrium:

Strategies: The principal's strategy σ_P is: for any history h_t, and proposal i, $\sigma_P(i|h_t) = 1$, i.e. at each history, the principal accepts any proposal with probability one. Every type of the agent proposes, at *any* history, her favorite available project with probability one. So, if $b \in S$, then, at any h_t, $\sigma_S(b|h_t) = 1$, and if $b \notin S$, then at any h_t, $\sigma_S(g|h_t) = 1$.

Beliefs: Fix any h_t (if $t = 0$, this is the null history). (i) If the agent is silent at

h_t, the beliefs are $\mu_M(h_t, \varnothing) = 1$ and $\mu_S(h_t, \varnothing) = 0$ for all $S \neq M$. (ii) If the agent proposes g at h_t, the beliefs are $\mu_G(h_t, g) = 1$ and $\mu_S(h_t, g) = 0$ for all $S \neq G$. (iii) If the agent proposes b at h_t, the beliefs are $\mu_B(h_t, b) = \frac{\mu_B}{\mu_B + \mu_M}$, and $\mu_M(h_t, b) = \frac{\mu_M}{\mu_B + \mu_M}$.

We now argue that the beliefs satisfy the requirements for PBE, and also satisfy Assumption 1. PBE says that beliefs need to be derived using Bayes' Rule *wherever possible*, so it imposes no restrictions on beliefs if at any h_t, the agent proposes i, such that $\sigma_S(i|h_t) = 0$ $\forall S$ such that $\mu_S(h_t) > 0$. So, for PBE, we only need to worry about the case where the agent's proposal at h_t has positive probability given this history. In this case, PBE requires the beliefs to be determined by Bayes' Rule. At any h_t, the only possible *on-path* proposals[12] are g and b (silence always is off-path at every history). g is a positive-probability proposal at h_t only if $\mu_G(h_t) > 0$, and in this case, the belief that we have specified, is precisely the one determined by Bayes' Rule. For b to be a positive probability proposal at h_t, we must have $\mu_B(h_t) + \mu_M(h_t) > 0$, so Bayes' Rule implies that $\mu_B(h_t, b) = \frac{\mu_B(h_t)}{\mu_B(h_t) + \mu_M(h_t)}$, and $\mu_M(h_t, b) = \frac{\mu_M(h_t)}{\mu_B(h_t) + \mu_M(h_t)}$. If $t = 0$, then we can obviously replace $\mu_M(h_t)$ and $\mu_B(h_t)$ with μ_M and μ_B in the above expressions. For $t > 0$, there are two possibilities. Either b was proposed at every $t' < t$ along this history. Then, we can show by induction, and Bayes' Rule that at every $t' < t$, $\mu_m(h'_t) = \frac{\mu_M}{\mu_B + \mu_M}$, and $\mu_B(h_t) = \frac{\mu_B}{\mu_B + \mu_M}$. The other case is that there exist periods along this history where b was not proposed. Then, let $t'' = \max\{t' < t | b \text{ was proposed at } t'\}$. Then, for every t such that $t'' < t \leqslant t$, b is proposed, and, we can again show by induction that beliefs derived from Bayes' Rule are indeed the ones we have specified.

Now, we come to the off-path proposals, i.e. if the agent proposes i, such that $\sigma_S(i|h_t) = 0$ $\forall S$ such that $\mu_S(h_t) > 0$. In this case, PBE imposes no restrictions. We only need to check if the beliefs satisfy Assumption 1. Silence is an off-path proposal at any history, and **bad** for every type, since if they proposed their favorite available project, it would be accepted. So, Assumption 1 imposes no restrictions. If $\mu_G(h_t) = 0$, and g is proposed, then, if $t > 0$, this proposal is **bad** for every type, so Assumption 1 imposes no restrictions. Same for an off-path proposal of b at $t > 0$.

Lemma 1. *The strategies and beliefs described above constitute a Perfect Bayesian Equilibrium.*

Proof. We have already argued that the beliefs that we described satisfy the require-

[12]The history h_t may itself be off-path. By on-path action we mean a proposal i such that there exists a type S such that $\mu_S(h_t) > 0$, and $\sigma_S(i|h_t) > 0$.

ments for being part of a PBE. Now, we only have to argue that in the continuation game starting at any history h_t, the strategies of the principal and agent constitute a Bayes Nash Equilibrium (BNE), given the principal's beliefs $\mu(h_t)$ at that history.

Fix any history h_t. If the agent is silent, there is no action for the principal to take. If the agent proposes g, then irrespective of $\mu(h_t)$, it is sequentially rational to accept g, since this is the highest payoff the principal can get. If the agent proposes b, again, the exact beliefs of the principal do not matter. Whatever the beliefs are, they have support in $\{S \subseteq N | b \in S\} = \{B, M\}$. So, the principal must believe with probability one that it is a type that *has* b, and that therefore will propose b in the next period, if the principal rejects this proposal (in fact, in every future period). So, it is sequentially rational for the principal to accept this proposal too.

For the agent, if she is of type G, it again has no profitable deviation to proposing g, as the principal will accept it if proposed. If she is of type B or M, again, the principal will accept b if proposed, no there is no profitable deviation to proposing b. □

Lemma 2. *The principal's payoff from the pooling equilibrium is the same as his payoff from the pooling mechanism.*

Proof. Along the equilibrium path, from type G, g is proposed and accepted at $t = 0$, and from types B and M, b is proposed and accepted at $t = 0$. This exactly replicates the implementation probabilities of $q_G(g) = 1$, and $q_B(b) = q_M(b) = 1$ from the pooling mechanism. □

A.2.2 Separating equilibrium:

Before describing the strategies and beliefs, we define two classes of histories.

Definition. *A* ***type 1*** *history h_t is one where there is* ***no*** *$t' < t^*(\delta)$ such that $h_t(t') = b$. In other words, there is no $t' < t^*(\delta)$ such that along h_t, b was proposed at t'. The null history is a* ***type 1*** *history.*

Definition. *A* ***type 2*** *history h_t is one where $t \neq 0$ and there* ***is*** *a $t' < t^*(\delta)$ such that $h_t(t') = b$. In other words, there is a $t' < t^*(\delta)$ such that along h_t, b was proposed at t'.*

Strategies:

- The principal's strategy σ_P is: At any history h_t, $\sigma_P(g|h_t) = 1$. So, g is accepted if proposed at any history. If h_t is such that $t < t^*(\delta)$, $\sigma_P(b|h_t) = 0$. If $t \geqslant t^*(\delta)$, and the history is of type 1, then $\sigma_P(b|h_t) = 1$. If $t \geqslant t^*(\delta)$, and the history is of type 2, then $\sigma_P(b|h_t) = 0$.

- If the agent's type is G, for any history h_t, $\sigma_G(g|h_t) = 1$.

- If the agent's type is B, and h_t is such that $t < t^*(\delta)$, then $\sigma_B(\varnothing|h_t) = 1$. If $t \geqslant t^*(\delta)$ and h_t is of type 1, then $\sigma_B(b|h_t) = 1$. If $t \geqslant t^*(\delta)$ and h_t is of type 2, then $\sigma_B(\varnothing|h_t) = 1$.

- If the agent's type is M, then at h_0, $\sigma(g|h_0) = 1$. Now consider h_t with $t > 0$. If h_t is of type 2, then $\sigma_M(g|h_t) = 1$. If h_t is of type 1, and $t < t^*(\delta)$, then $\sigma_M(\varnothing|h_t) = 1$. If h_t is of type 1, and $t \geqslant t^*(\delta)$, then $\sigma_M(b|h_t) = 1$.

Beliefs:
At the null history h_0:

- If the agent is silent, the beliefs are $\mu_B(h_0, \varnothing) = 1$. If the agent proposes g, the beliefs are $\mu_G(h_0, g) = \frac{\mu_G}{\mu_G+\mu_M}$, $\mu_M(h_0, g) = \frac{\mu_M}{\mu_G+\mu_M}$. If the agent proposes b, the beliefs are $\mu_M(h_0, b) = 1$.

- These beliefs all satisfy the conditions for PBE and Assumption 1.The agent staying silent and proposing g both occur with positive probability at h_0, as $\sigma_B(\varnothing|h_0) = 1$ and $\sigma_G(g|h_0) = \sigma_M(g|h_0) = 1$, and these two cases, beliefs are the ones determined by Bayes' Rule. Proposing b is an off-path action here, thus, beliefs can be arbitrary as per PBE. We already argued in the main text that they satisfy Assumption 1.

At a type 1 history h_t**, where** $t > 0$:

- If the agent is silent, beliefs are $\mu_B(h_t, \varnothing) = 1$. If the agent proposes g, beliefs are $\mu_G(h_t, g) = 1$. If the agent proposes b, and $t < t^*(\delta)$ beliefs are $\mu_M(h_t, b) = 1$. If $t \geqslant t^*(\delta)$, beliefs are $\mu_B(h_t, b) = 1$.

- The beliefs are consistent with PBE and satisfy Assumption 1. To see this, observe that silence is a positive probability action at h_t only if $t < t^*(\delta)$ and $\mu_B(h_t) > 0$ or $\mu_M(h_t) > 0$. This is the case only if $h_t(t-1) = \varnothing$, and in this case beliefs are the same as the ones determined by Bayes' Rule. Proposing g is never a positive probability action at h_t so beliefs can be arbitrary in a PBE. Moreover, proposing g at $t > 0$, at a type 1 history, is a **bad** deviation for both the mixed and the good types (i.e. the types that *can* propose g), so Assumption 1 imposes no restrictions and the belief $\mu_G(h_t, g) = 1$ satisfies Assumption 1.

- Proposing b when $t < t^*(\delta)$ is also off-path, and PBE imposes no restrictions. In this case, as we argued in the main text, Assumption **??** imposes no restrictions because this deviation is **not bad** for both mixed and bad types.

  Proposing b is a positive probability action only if $t \geqslant t^*(\delta)$. (1)If $t = t^*(\delta)$, this is the case only if $h_t(t-1) = \varnothing$. In this case, $\mu_B(h_t) = 1$, and $\sigma_B(b|h_t) = 1$, so beliefs are precisely the ones determined by Bayes' Rule. (2) If $t > t^*(\delta)$, then proposing b is a positive probability event only if $h_t(t-1) = \varnothing$ or $h_t(t-1) = b$. In this case, $\mu_B(h_t) = 1$ and $\sigma_B(b|h_t) = 1$, so again, the beliefs we specified are the ones determined by Bayes' Rule. If b is proposed at $t \geqslant t^*(\delta)$, and it is a zero-probability proposal, then, Assumption 1 imposes restrictions only if it is the first off-path action along h_t. This cannot be the case when $t > t^*(\delta)$. If $t = t^*(\delta)$, b is off-path only when $\mu_B(h_t) = 0$, and this cannot be the case at a type 1 history where b is the first off-path action at $t^*(\delta)$.

At a type 2 history:

- If the agent is silent, the beliefs are $\mu_B(h_t, \varnothing) = 1$. If the agent proposes g, beliefs are $\mu_M(h_t, g) = 1$. If the agent proposes b, beliefs are $\mu_M(h_t, b) = 1$.

- Note that Assumption 1 imposes no restrictions following a probability zero action at a type 2 history because this would not be the first off-path action along this history; b has already been proposed before $t^*(\delta)$ at least once. So we only need to check if the beliefs are consistent with PBE. If the agent is silent, there are two possibilities. Either, $h_t(t-1) = \varnothing$, in which case $\mu_B(h_t) = 1$, so the beliefs that we mentioned are precisely the ones determined by Bayes' Rule. If $h_t(t-1) = b$, or $h_t(t-1) = g$, silence is a probability zero action at h_t, and PBE allows beliefs to be arbitrary.

- If the agent proposed b at h_t, irrespective of what was proposed at $t-1$, b is a probability zero action. This is because $\sigma_M(b|h_t) = \sigma_B(b|h_t) = 0$ at h_t. So again, PBE allows beliefs to be arbitrary. If agent proposed g at h_t, g is a positive probability action here only if $h_t(t-1) = b$, or $h_t(t-1) = g$. In this case since $\mu_M(h_t) = 1$ and $\sigma_M(g|h_t) = 1$, beliefs are precisely the ones determined by Bayes' Rule. If $h_t(t-1) = \varnothing$, beliefs can be arbitrary.

Lemma 3. *The strategies and beliefs described above constitute a Perfect Bayesian Equilibrium.*

Proof. We have already argued that the beliefs that we described satisfy the requirements for being part of a PBE and satisfy assumption 1. We must now argue that in the continuation game starting at any history h_t, the strategies of the principal and agent constitute a Bayes Nash Equilibrium (BNE), given the principal's beliefs $\mu(h_t)$ at that history.

For the principal, fix any history h_t. If the agent is silent, there is no action for the principal to take. If the agent proposes g, then irrespective of $\mu(h_t)$, it is sequentially rational to accept g, since this is the highest payoff the principal can get. If the agent proposes b, we need to consider three cases. (i) $t < t^*(\delta)$: The principal's strategy is to reject b at such a history. His beliefs following a proposal of b at h_t are $\mu_M(h_t, b) = 1$. If the principal accepts, he gets α_b, and if he rejects, he expects the agent to propose g in the next period. So, rejection is indeed sequentially rational if the principal is sufficiently patient. (ii) $t \geqslant t^*(\delta)$ and the history is of type 1. In this case, if b is proposed, the principal's beliefs are $\mu_B(h_t, b) = 1$ and if he rejects b, he expects b to be proposed again in the next period. So, accepting b is sequentially rational. (iii) $t \geqslant t^*(\delta)$ and the history is of type 2. In this case, the principal's strategy is to reject b. His beliefs following a proposal of b are $\mu_M(h_t, b) = 1$, so the same reasoning as case (i) follows, and rejection is sequentially rational.

For the agent, fix a history h_t. If her type is G, her strategy is to propose g, which is optimal, since the principal would accept it. If her type is B, and (i) $t < t^*(\delta)$, her strategy is to stay silent. Consider a one-shot deviation where she deviates by proposing b. This proposal would be rejected and since her strategy involves staying silent at any future period following this rejection, therefore getting a payoff of zero, she cannot be better off by this deviation. (ii) If $t \geqslant t^*(\delta)$ and the history is of type 1, proposing b is optimal since it would be accepted and α_b is the highest payoff the agent can get. (iii)If $t \geqslant t^*(\delta)$ and the history is of type 2, silence is optimal. Consider a one-shot deviation where the agent proposes b instead. It would be rejected, and the agent's strategy is to stay silent in each period that follows. So, this is not a profitable deviation.

If the agent's type is M, at h_0, her strategy is to propose g. Consider the one-shot deviation where she proposes b instead. It will be rejected and she will propose g at $t = 1$, which will get accepted. Clearly, this is not profitable, as she can propose g at $t = 0$ and it will get accepted. If the one-shot deviation involves silence at $t = 0$, her strategy then is to stay silent till $t^*(\delta)$, at which point she proposes b can it is accepted. So her payoff is $\delta^{t^*(\delta)}\alpha_b$ which is $\leqslant \alpha_g$. So, this deviation is not profitable either. We can similarly rule out deviations at other histories.

□

A.3 Comparison with Static Delegation Game

The dominant approach in the literature, starting with Armstrong & Vickers (2010), has been to model the project selection problem as a *one shot* interaction between the principal and the agent, that takes the form of **constrained delegation**. The principal **commits** to a *permission set* $A \subseteq N$, which is the set of projects that he will approve, if proposed. An agent of type S then either chooses a feasible project from this set, or, if $S \cap A = \varnothing$, the permission set it such that it does not contain *any* feasible project, so she chooses nothing, and no project is implemented.

The choice for the principal here boils down to whether or not he should allow the bad project. Observe that here, there is no opportunity for the principal to *learn* about the agent's type, and base her decision on the information she has elicited from the agent. Therefore, it should not come as a surprise that if the principal can commit to a strategy in our dynamic elicitation game, he can do at least as well (and sometimes strictly better) than this static delegation game. This is because by committing to a strategy in the dynamic game, the principal can commit to his decision as a *function* of what the agent proposes, and can therefore screen different types of the agent more effectively. Additionally, since in the dynamic game, the commitment payoff can always be approximated in equilibrium as $\delta \to 1$, therefore, the principal-optimal equilibrium of the dynamic game is at least as good as the equilibrium of this static delegation game.

In particular, whenever the separating equilibrium is optimal in the dynamic game, the principal's payoff from the separating equilibrium is strictly higher than his payoff from the static delegation game. The main idea behind this is that the outcome of the separating equilibrium from our game cannot replicated here. In the separating equilibrium, conditional on the agent being of bad type, the bad project is implemented with a delay. But in the static game, if the principal wants the mixed type to choose g, he must *not* allow b, and therefore nothing is implemented conditional on the agent's type being bad. The proposition below, the proof of which is in the Appendix, says exactly this. Recall that λ^* is the likelihood ratio of the mixed and bad types above which the separating equilibrium is optimal in the dynamic game.

Proposition 4. *The principal always does weakly better in an equilibrium of the dynamic game than in the static delegation game. She does strictly better with the dynamic game if and only if $\lambda > \lambda^*$.*

Proof. First, we solve for the equilibrium of the static delegation game.[13]. An equilibrium here is characterized by the principal's optimal permission set $A^*(\mu_0)$, where μ_0 is the prior.

Claim 4. *Let $\lambda^{**} := \frac{\pi_b}{\pi_g - \pi_b}$. Then, $A^* = \{b, g\}$ if and only if $\lambda < \lambda^{**}$. Otherwise, $A^* = \{g\}$.*

Proof. It must be that $g \in A^*(\mu_0)$. The principal must decide whether or not to include b in A^*. If she includes b, both mixed and bad types would choose b, and only the good type would choose g. So, the principal's payoff from setting $A = \{b, g\}$ is:

$$\pi(\{b, g\}) := \mu_0(G)\pi_g + \mu_0(M)\pi_b + \mu_0(B)\pi_b$$

Her payoff from not allowing b is:

$$\pi(\{g\}) := \mu_0(G)\pi_g + \mu_0(M)\pi_g$$

This is because if b is not included in the permission set, then the bad type has no feasible project that's allowed and chooses nothing. Comparing $\pi(\{b, g\})$ and $\pi(\{g\})$, we have that $\pi(\{b, g\}) \leqslant \pi(\{g\})$ if and only if $\frac{\mu_0(M)}{\mu_0(B)} \geqslant \frac{\pi_b}{\pi_g - \pi_b}$.

□

[13]For a more general treatment of the constrained static delegation problem, see Armstrong & Vickers (2010). For completeness, we provide a characterization of the optimal permission set for our setting.

Observe that $\lambda^{**} > \lambda^*$, which was the threshold of $\frac{\mu_0(M)}{\mu_0(B)}$ above which the separating equilibrium was optimal and below which, the pooling equilibrium was optimal in the dynamic game. We now consider three case, and show that in all three cases, the principal-optimal equilibrium in the dynamic game does weakly better than the optimal static delegation, and strictly better in two out of three cases.

Case 1 *:* $\lambda \leqslant \lambda^*$. In this case, in the dynamic game, the pooling equilibrium is optimal, and in the static delegation game, $A^* = \{b, g\}$. In both cases, the principal gets the same payoff because in the dynamic game, all types choose their favorite projects at $t = 0$.

Case 2*:* $\lambda^* < \lambda \leqslant \lambda^{**}$. In this case, in the dynamic game the separating equilibrium is optimal. So, principal's payoff in the dynamic game is $\pi^*_{sep} := \mu_0(G)\pi_g + \mu_0(\pi_M)\pi_b + \mu_0(B)\delta^{t^*(\delta)}\pi_b$. In the static delegation game, however, permitting b is optimal, so the principal's payoff is $\pi^*_{del} := \mu_0(G)\pi_g + \mu_0(M)\pi_b + \mu_0(B)\pi_b$. Since $\lambda > \lambda^*$, we have that $\pi^*_{sep} > \pi^*_{del}$.

Case 3*:* $\lambda > \lambda^{**}$: In this case, the separating equilibrium is still optimal in the dynamic game. In the static delegation game however, now, $a^* = \{g\}$, so the principal's payoff is $\pi^*_{del} = \mu_0(G)\pi_g + \mu_0(M)\pi_g$, which is obviously lower than her payoff from the separating equilibrium.

This completes the proof of the Proposition.

□

A.4 Other Extensive Forms

An extensive form should specify who the proposer will be at any history, and what they are allowed to propose. Formally, at any history h_t, the extensive form specifies the proposer, $P(h_t)$ and the set of permissible offers, $\mathcal{O}(h_t) \subseteq 2^{\mathcal{N}}$, so any offer $O(h_t)$ is a subset of $\mathcal{N}$.[14] When an offer is made by the proposer, the other party responds by accepting a project in the offer or rejecting the offer altogether. A history is a

[14]An offer can be $\varnothing$, this can be interpreted as the party making the offer permitting no project to be chosen by the other, like the agent choosing to stay silent in our dynamic game. If the offer is $\varnothing$, then the party responding to the offer has no option but to reject it, as there is no project to accept.

sequence of offers that have been rejected. [15]

The set of actions feasible for the agent at any history is type-dependent; if he is the proposer, he can only include available projects in her offer, and if the principal is the proposer, he can only accept an available project. Formally, if, at h_t, the proposer is the agent, and the agent's type is S, then we must have $O(h_t) \subseteq S$. On the other hand, if the proposer is the principal, the offer $O(h_t)$ can be any subset of $\mathcal{N}$ that's in $\mathcal{O}(h_t)$, but the agent can only accept a project in $O(h_t) \cap S$.

An extensive form here is therefore simply a dynamic game with type-dependent action space for the agent at any history. If any project from an offer is accepted, the game ends, and players get their discounted payoffs; otherwise the game proceeds to the next period. Our equilibrium concept is Perfect Bayesian Equilibrium; both players play sequentially rationally and the principal's beliefs about the agent's type are updated according to Bayes' rule whenever possible.

Fix an extensive form and also a commitment strategy of the principal and the agent's best response to it. An outcome here is still (i, t), which means that project i is implemented at time t. As in the proof of Proposition 2, a commitment strategy of the principal, and the agent's best response together induce a probability distribution over outcomes, and we can then construct a mechanism in the class of static stochastic mechanisms that we defined in Section A.1, such that the principal's payoff from her commitment strategy is achieved. So, this class of mechanisms represents an upper bound to what the principal can achieve with commitment, and therefore in an equilibrium of any of these extensive forms. Since the payoff from the optimal mechanism can already be achieved in an equilibrium of our dynamic game, it is clear that none of these extensive forms can do better, from the principal's point of view.

A.5 Refinement

In this section, we provide a more general description of our refinement. Recall that in the main text, we only dealt with a particular kind of deviation: One where the agent either makes an off-path proposal of b at $t = 0$, or is silent till $t' - 1$, and makes an off-path proposal of b at t', such that b is her **first** off-path proposal, and the silent at every $t < t'$ is on-path along this history. Of course, the agent's on-path behavior prior to taking her first off-path action can involve something other than staying silent.

[15] Tying this definition back to our game, it is the protocol where at any h_t, $P(h_t)$ is the agent, and the set of permissible offers is $\mathcal{O}(h_t) = \mathcal{O} = \{\{i\} | i \in \mathcal{N}\} \cup \varnothing$, i.e. the singleton subsets of $\mathcal{N}$.

Also, her first off-path action does not need to be a proposal of b: it can be staying silent at a history where she was supposed to propose something. It could also be an off-path proposal of g.

Suppose the first off-path proposal i_{dev} is made at time t', at history $h_{t'}$. So, along $h_{t'}$, at all $t < t'$, the agent's proposal i_t is on-path at $h_t(h_{t'})$, which is the history obtained by truncating $h_{t'}$ at t. For each type S, we want to compute the maximum payoff from this deviation for type S, and to do so, we divide the agent's payoff into two parts: Her expected payoff from her on-path proposals *before* the first off-path proposal, and her payoff following the first off-path proposal. Since the agent's proposals are on-path till $t' - 1$, the principal responds to these according to his equilibrium strategy σ_P^*. We denote the expected payoff for type S till t by $\alpha_S^*(< h_{t'})$. To highlight a couple of important things about $\alpha^{*_S}(< h_{t'})$, we first give the example of a mixed strategy equilibrium.

Example 1. *Suppose* $\frac{\pi_b}{\delta \pi_g} \leqslant \frac{\mu_0(M)}{\mu_0(B)+\mu_0(M)}$. *Then, there exists a partially separating, mixed strategy equilibrium with the following on-path behavior:*

- *The principal accepts g at any history. If b is proposed at $t = 0$, he accepts it with probability* $y^* = \frac{(1-\delta)\alpha_g}{\alpha_b - \delta\alpha_g}$.[16] *If b is proposed at $t > 0$, the principal rejects it irrespective of history.*

- *The good type agent proposes g with probability one at $t = 0$. The bad type agent proposes b with probability one at $t = 0$, and stays silent in every subsequent period. The mixed type agent randomizes between proposing g and proposing b at $t = 0$. She proposes b at $t = 0$ with probability* $x^* = \frac{\mu_0(B)c}{\mu_0(M)(1-c)}$, *where* $c = \frac{\mu_0(B)}{\delta\mu_0(M)}$. *If b is rejected at $t = 0$, she proposes g with probability one at $t = 1$. Note that the condition* $\frac{\pi_b}{\delta\pi_g} \leqslant \frac{\mu_0(M)}{\mu_0(B)+\mu_0(M)}$ *ensures that $x^* \leqslant 1$.*

In the above equilibrium, it is easy to see that given the agent's strategy, the principal is indifferent between accepting and rejecting b at $t = 0$. If b is proposed at $t > 0$, the principal's off-path belief is that its the mixed type with probability one, and will propose g in the next period. It is also easy to verify that given the principal's strategy, the mixed type is indifferent between proposing g at $t = 0$, and getting α_g, or proposing b at $t = 0$, and then proposing g if rejected, and getting an expected payoff of $y^*\alpha_b + (1 - y^*)\delta\alpha_g$.

[16] $y^* < 1$ as $\alpha_b > \alpha_g$.

We now make some observations about $\alpha_S^*(< h_{t'})$:

- For any type S, $\alpha_S^*(< h_{t'})$ need not correspond to the expected payoff that type S gets till time t' from following her own equilibrium strategy. Even though $h_{t'}$ is an on-path history, it may not arise with positive probability if S follows σ_S^*. It could be that S mimics the strategy of another type till time t', before making her first off-path proposal. For example, in the example of the mixed strategy equilibrium given above, the mixed type could deviate by proposing b at $t = 0$ with probability one, then staying silent till some $t' - 1 > 0$, and proposing b at t'. Till the off-path proposal of b at t', the mixed type's behavior in this deviation is same as the on-path behavior of the bad type.

- Suppose before her first off path proposal i_{dev} at $h_{t'}$, the agent was **not** silent at every $t' < t'$. The, even though along this history, the on-path proposals before t' were rejected (and therefore $h_{t'}$ was reached), the ex ante expected payoff till period t' from these proposals need not be zero. Again, consider the equilibrium in Example 1. Suppose the agent proposes b at $t = 0$, then stays silent till $t' - 1 > 0$, and proposes b at t'. Then, the ex ante expected payoff before the off-path proposal of b at t', is $y^*\alpha_b$.

Following the first off-path proposal, we still consider, for any type S of the agent, the highest PBE payoff that this type can get in the continuation game following the deviation, for *any* beliefs of the principal following the deviation, and denote this highest payoff by $\alpha^*(h_{t'}, i_{dev})$. The interesting cases are when $i_{dev} \in \{\varnothing, b\}$, and $b \in S$.[17] Recall that the highest PBE payoff for any type is the one from the *pooling equilibrium*, and this equilibrium exists for any beliefs of the principal following the history $(h_{t'}, i_{dev})$, thus, if $i_{dev} = b$, then $\alpha^*(h_{t'}, b) = \alpha_b$. If $i_{dev} = \varnothing$ and $b \in S$, then $\alpha^*(h_{t'}, b) = \delta\alpha_b$, i.e. the highest payoff following off-path silence is the payoff from the pooling equilibrium being played next period onward.

We are finally ready to state our restriction more generally.

Definition 2. *A deviation i_{dev} at history $h_{t'}$ is **bad** for type S if her maximum payoff from the deviation, $\alpha_S^*(< h_{t'}) + \delta^{t'}\alpha^*(h_{t'}, i_{dev})$ is strictly lower than $\alpha_S^*(\sigma_S^*)$, where $\alpha_S^*(\sigma_S^*)$ is the agent's payoff from following her equilibrium strategy.*

[17] If $i_{dev} = g$,then this proposal is obviously accepted by the principal, even if its off-path, so in this case, $\alpha^*(h_{t'}, g) = \alpha_g$.

Assumption 5. *Suppose i_{dev} at history $h_{t'}$ is the first off-path proposal along $h_{t'}$. Then, if this deviation is **bad** for some type S, and **not bad** for another type S', then, at history $(h_{t'}, i_{dev})$, the principal's off-path beliefs attach probability zero to type S.*

Claim 5. *The off-path beliefs in the mixed strategy equilibrium in Example 1 do not satisfy Assumption 5.*

Proof. Suppose that the agent proposes b at $t = 0$, is silent until t' if this proposal is rejected, and proposes b at t'. Then b is the first off-path proposal along h_t, and $\alpha^*_S(h_{t'}, b) = y^*\alpha_b + (1-y^*)\delta^{t'}\alpha_b$. This is the highest payoff from this deviation because the proposal of b at $t = 0$ is on-path, so the principal accepts with probability y^*, according to his equilibrium strategy. If this is rejected the agent is silent till t', which is on-path because it mimics the equilibrium behavior of the bad type. At t', the first off-path proposal of b is made, and the highest PBE payoff following this proposal is if its accepted.

Suppose t' is large enough that $y^*\alpha_b + (1 - y^*)\delta^{t'}\alpha_b < y^*\alpha_b + (1 - y^*)\delta\alpha_g$, so the mixed type would strictly prefer her equilibrium payoff to the highest payoff from this deviation. So, this deviation is **bad** for the mixed type. It is **not bad** for the bad type though. So, the off-path beliefs following this deviation should attach probability one to the bad type. However, in the equilibrium we described, the beliefs attach probability one to the mixed type following the deviation.

□

A.6 Proof of Theorem 2

Proof. The proof proceeds in three steps. Broadly, we first show that in solving for the optimal static stochastic mechanism, where the class of mechanisms that we restrict attention to is identical to the one that we considered for the two project case, the optimization problem can be divided into two parts. Then we show that if the solution to the second part corresponds to a value in a certain set, the payoff from the optimal mechanism can be replicated in equilibrium. Lastly, we show that the solution to the second part must always correspond to a value in this set. Since commitment to a strategy in the dynamic game can never lead to a higher payoff that commitment to a static stochastic mechanism, this would complete the proof.

Let the set of types be $\{S_1, S_2 \dots S_M\}$ where for any $i < i'$, we have that $S_{i'} \subset S_i$. For any type S_i, let $\mu(S_i) = \mu_i$ and let $q_{i,j}$ be the probability of implementation of

$j \in S_i$ in the mechanism when the report is S_i. Let the expected value to the agent, corresponding to any report S_i, be denoted by E_i, where $E_i := \sum_{j \in S_i} q_{i,j}\alpha_j$. Observe that due to type-dependent message spaces and the support of $\mu(.)$, the IC constraints here boil down to $(M-1)$ inequalities:

$$E_1 \geqslant E_2, \ldots \geqslant E_M,$$

and we refer to the inequalities $E_{i+1} \ldots \geqslant E_M$ as the IC constraints *below* i and the inequalities $E_1 \geqslant E_2 \ldots \geqslant E_i$ as the IC constraints *above* i.

Lemma 4. *In any optimal mechanism, any report must generate the same expected payoff for the agent. Formally, for any two reports S_i and $S_{i'}$, it must be that $E_i = E_{i'}$.*

Proof. We prove this by contradiction. Suppose, in an optimal mechanism there are i, i' such that $E_i \neq E_{i'}$. Without loss, let $i < i'$. This implies, given the nature of the support of $\mu(.)$, that $S_{i'} \subset S_i$. Since the optimal mechanism is IC, it must be that $E_i \geqslant E_{i'}$, and since $E_i \neq E_{i'}$, we have that $E_i > E_{i'}$. This in turn implies that we can find *consecutive types* $k, k+1$ such that $i \leqslant k < k+1 \leqslant i'$ and $E_k > E_{k+1}$. So without loss, let $i' = i+1$. Let i^* be the lowest indexed project in S_i.

- **Case 1**: *Corresponding to report S_i, $q_{i,j} > 0$, for some $j > i^*$.*

  In this case, consider the following perturbation: Let $q'_{i,j} = q_{i,j} - \varepsilon$ and $q'_{i,i^*} = q_{i,i^*} + \varepsilon$. Now, $E'_i = E_i - \varepsilon(\alpha_j - \alpha_i) < E_i$. The ε in the perturbation is small enough that $E'_i > E_{i+1}$. Other than this change, all allocation probabilities corresponding to all other reports are unchanged, relative to the original mechanism. This new mechanism is IC, because since the original mechanism was IC, and we have reduced E_i, all IC constrains *above* i still hold. The inequality between E'_i and E_{i+1} is preserved, so this IC still holds. All IC constraints *below* i still hold, clearly. Thus we have constructed another IC mechanism in $(*)$ that gives a strictly higher expected payoff to the principal, as his payoff from type S_i increases by $\varepsilon(\pi_{i*} - \pi_j)$. Thus, the mechanism we started out with cannot be optimal.

- **Case 2**: *Corresponding to report S_i, $q_{i,j} = 0$, for every $j > i^*$.*

  We can again construct an IC mechanism in $(*)$ that gives strictly higher expected payoff to the principal. Since $q_{i,j} = 0$, for every $j > i^*$, we have that $E_i = q_{i,i^*}\alpha_{i^*} \leqslant \alpha_{i^*}$, as only i^* might have positive allocation probability in S_i. Also, $E_i > E_{i+1}$, so it must be that $\sum_j q_{i+1,j} < 1$. This is because all projects in S_{i+1} have weakly higher payoff for the agent than α_{i^*}, so if their allocation probabilities sum up to one, we would have $E_{i+1} \geqslant \alpha_{i^*} \geqslant E_i$, which cannot be. So, since $\sum_j q_{i+1,j} < 1$, in particular, $q_{i+1,(i+1)^*} < 1$. Consider the following perturbation: let $q'_{i+1,(i+1)^*} = q_{i+1,(i+1)^*} + \varepsilon$ where ε is small enough that $E'_{i+1} = E_{i+1} + \varepsilon\alpha_{(i+1)^*} < E_i$, so the IC constraint between $i, i+1$ is preserved. Everything else is unchanged with respect to the original mechanism. Clearly, all IC constraints *above* i hold, and all *below* i hold as well as we increased E_{i+1}. This mechanism gives the principal a higher expected payoff since the payoff from type S_{i+1} has increased. So, the mechanism we started out with cannot be optimal.

This completes the proof of our claim that in any optimal mechanism, *any* report must generate the same expected payoff for the agent.

□

Now that we have shown this, the principal's optimization problem (finding the payoff-maximizing mechanism among all IC mechanisms in $(*)$) can be divided into two parts. First, for any expected value v, find the optimal mechanism corresponding to *this* v; the mechanism that maximizes the principal's payoff when each report generates an expected payoff of v for the agent. Then, maximize the principal's payoff over the possible values of v, i.e. find the values of v the optimal mechanism corresponding to which generates the highest expected payoff for the principal. Our aim is not to solve for the optimal mechanism, but rather show it is always the case that the principal's expected payoff from the optimal mechanism can be attained in equilibrium. We do this in the steps that follow.

Lemma 5. *In any optimal mechanism, it cannot be that $v > \alpha_{M^*}$, where M^* is the lowest indexed project in S_M.*

Proof. Suppose in the optimal mechanism, $v > \alpha_{M^*}$. Then it must be that there is a project $j \in S_M$ such that $j > M^*$, since the expected value that type S_M gets is $> \alpha_{M^*}$. So, we can perturb this mechanism as follows:$q'_{M,j} = q_{M,j} - \varepsilon$, and $q'_{M,M^*} = q_{M,M^*} + \varepsilon$. Clearly, the new mechanism is IC as there is no IC below M. And it results in higher expected payoff for the principal. □

Lemma 6. *Let the set of all projects in types $\{S_1, S_2 \dots S_M\}$ be $\{\alpha_1, \alpha_2, \dots, \alpha_N\}$, where $\alpha_1 < \alpha_2 \dots \alpha_N$. For any $v \in \{\alpha_1, \alpha_2, \dots, \alpha_N\}$, such that $v \leqslant \alpha_{M*}$, there exists an equilibrium of the dynamic elicitation game in which the principal attains the payoff from the optimal mechanism corresponding to v, as $\delta \to 1$.*

Proof. Recall that for any project i, $\frac{\pi_1 - \pi_i}{\alpha_i - \alpha_1} = K$ where K is a constant. Let $v = \alpha_k < \alpha_{M*}$. For each S_i, we solve the following problem:

$$\begin{aligned} \max_{\{q_{i,j} | j \in S_i\}} \qquad & \sum_{j \in S_i} q_{i,j} \pi_j \\ \text{subject to} \qquad & \sum_{j \in S_i} q_{i,j} \alpha_j = \alpha_k \end{aligned} \tag{1}$$

There are two possibilities: Either $i^* < k$ or $i^* \geqslant k$. If $i \geqslant k$, the solution is $q^*_{i,i*} = \frac{\alpha_k}{\alpha_i}$, because we cannot do better than assigning positive probability to *only* the principal-favorite project in S_i, which is i^*, and since $i^* \geqslant k \implies \alpha_i \geqslant \alpha_k$, we can do so.

Now, let us consider the case where $i < k$. Here, $\alpha_{i*} < \alpha_k$, so we can no longer assign positive probability only to i^* in S_i. In this case, any solution to the above optimization problem must satisfy $\sum_j \{q^*_{i,j} | j \in S_i\} = 1$. If $\sum \{q^*_{i,j} | j \in S_i\} < 1$, then in particular $q^*_{i,i*} < 1$, and $q^*_{i,j} > 0$ for *some* $j' > i^*$, because we must have $\sum_{j \in S_i} q^*_{i,j} \alpha_j = \alpha_k$. Fix any such $j' > i^*$, such that $q^*_{i,j'} > 0$. We can now perturb the allocation probabilities as follows: Let $q^{**}_{i,j'} = q^*_{i,j'} - \varepsilon$, $q^{**}_{i,i*} = q^*_{i,i*} + \varepsilon \frac{\alpha_j}{\alpha^*_i}$, and $q^{**}_{i,j''} = q^*_{i,j''}\ \forall\ j'' \neq \{i^*, j\}$. It is straightforward to check that $\sum_{j \in S_i} q^{**}_{i,j} \alpha_j = v$, the principal's expected payoff is strictly higher, and for ε small enough, $\sum_{j \in S_i} q^{**}_{i,j} \alpha_j \leqslant 1$. So, if $i < k$, we must have $\sum \{q^*_{i,j} | j \in S_i\} = 1$. The constraint in the optimization problem can be thus be rewritten substituting $q_{i,i**} = 1 - \sum_{\{j \in S_i | j < i^{**}\}} q_{i,j}$, where i^{**} is the highest indexed project in S_i.

$$\sum_{\{j \in S_i | j < i^{**}\}} q_{i,j} (\alpha_{i**} - \alpha_j) = \alpha_{i**} - \alpha_k \tag{2}$$

We can also, after the same substitutions, rewrite the objective function and get:

$$\pi_{i^{**}} + \sum_{\{j \in S_i | j < i^{**}\}} q_{i,j}(\pi_j - \pi_{i^{**}}),$$

which, after substituting (2), is just equal to

$$\pi_{i^{**}} + \sum_{\{j \in S_i | j < i^{**}\}} q_{i,j} K(\alpha_{i^{**}} - \alpha_j) = \pi_{i^{**}} + (\alpha_{i^{**}} - \alpha_k)K$$

Observe that the last expression is a constant independent of allocation probabilities. So, *any* allocation probabilities that satisfy $\sum\{q_{i,j} | j \in S_i\} = 1$ and $\sum_{j \in S_i} q_{i,j}\alpha_j = \alpha_k$, solves the optimization problem. In particular, $q_{i,k} = 1$ solves (1) when $i < k$.

To sum up, for any $v \in \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, *an* optimal mechanism *corresponding to* v is as follows:

$$q^*_{i,i^*} = \frac{\alpha_k}{\alpha^*_i}, q^*_{i,j} = 0 \; \forall \; j > i^*, \; if \; i^* > k,$$

and

$$q^*_{i,k} = 1, q^*_{i,j} = 0 \; \forall \; j \neq k, \; if \; i^* \leqslant k$$

We now construct an equilibrium that replicates the payoff from the above mechanism. The construction is very similar to our *separating* equilibrium from the two-project case. Fix $v = \alpha_k$. Consider the equilibrium where on path, at $t = 0$, all types that have project k report it, and this proposal is accepted right away. For every S_i such that $i^* > k$, there exists a threshold $t^*_{i*}(\delta)$, such that type S_i, which does not have project k, proposes i at t^*_i, which is then accepted by the principal. We define this threshold inductively:

$$t^*_{k+1}(\delta) := \min\{t : \alpha_k \geqslant \delta^t \alpha_{k+1}\},$$

and, given that we have defined t^*_{k+j}, we define t^*_{k+j+1} as follows:

$$t^*_{k+j+1}(\delta) := \min\{t : \delta^{t^*_{k+j}(\delta)}\alpha_{k+j} \geqslant \delta^t \alpha_{k+j+1}\}$$

We omit the details of the strategies, as they are very similar to the *separating* equilibrium. But intuitively, this on path behavior can be supported in equilibrium as if the principal sees a proposal $i^* > k$ before t^*_{i*}, his off path belief is that it is type S_1 with probability one, and if this proposal is rejected, S_1 will propose 1 in the next period. The thresholds are such that any type that does not have k will find it optimal to propose the principal's favorite project that it has, at the appropriate threshold.

Note that in this proof we have implicitly assumed that all types where $i^* < k$ have project k. In case they do not, this construction does not work. However, the next Lemma will show that we do not have to worry about these cases; if such an $\alpha_k = v$ in the optimal mechanism, we can find another v' that attains same or strictly higher payoff for the principal, such that the optimal mechanism corresponding to this v' is implementable in equilibrium.

□

We have thus shown that every optimal mechanism must have *some* v which is the expected payoff to each type of the agent, and if the optimal mechanism has $v^* \in \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$, there always exists an equilibrium where the principal attains the payoff from the optimal mechanism as $\delta \to 1$. We now show, in the next lemma, that there is always a $v \in \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ such that the optimal mechanism *corresponding to* v is indeed optimal. This would complete our proof that commitment payoff can be attained for this case of nested types.

Lemma 7. *For any v such that $v \in (\alpha_k, \alpha_{k+1})$ for some $k \in \{1, 2 \ldots N-1\}$, either v cannot be part of the optimal mechanism, or there exists $v' \in \{\alpha_1, \alpha_2 \ldots \alpha_N\}$ such that the principal's payoff from the optimal mechanism corresponding to v' is the same as his payoff from the optimal mechanism corresponding to v.*[18]

Proof. Let $v \in (\alpha_k, \alpha_{k+1})$ for some $k \in \{1, \ldots, N\}$. The principal's objective is to maximize his expected payoff by choosing implementation probabilities for each type:

$$\max \sum_{j \geqslant i*} q_{i,j}\pi_j \quad \text{subject to} \quad \sum_{j \geqslant i*} q_{i,j}\alpha_j = v, \quad \forall i \in \{1, \ldots, M\}$$

For all types S_i with $i^* \geqslant k+1$, the optimal mechanism assigns $q_{i,i*} = \frac{v}{\alpha_{i*}} < 1$ and $q_{i,j} = 0$ for all other $j > i^*$.

For the rest of the types S_i with $i \leqslant k+1$, as we argued in Lemma 6, we have $\sum_j q_{i,j} = 1$. Since given the constraint that the agent's expected payoff equals v, any randomization is optimal, we consider one particular randomization as part of the optimal mechanism.

[18]The perturbations that we construct here will also work if $v = \alpha_k$ sor some k butt all types where $i^* < k$ do not have k.

An optimal mechanism that corresponds to the expected value $v \in (\alpha_k, \alpha_{k+1})$ is as follows:

- Consider $i = \min\{i'|i'^* < k+1\}$. In this case, $(i+1)^*$ is the lowest indexed project of the type S_{i+1}, so it must be that $(i+1)^* \geqslant k+1$. Also all types above $(i+1)$ will have both i^* and i^*+1.
- for all types above i, only projects i^* and i^*+1 are implemented with positive probability, with the appropriate mixture to provide the expected payoff of v, Let these probabilities be q_{i^*} and $q_{(i+1)^*}$.
- for all types below i, only the project with the lowest index is implemented with positive probability.

We now argue that there exists some v' such that there is an IC mechanism in which each type of the agent gets v' and the principal gets a strictly higher payoff than the mechanism we describe above. Consider the following perturbation:

- for all types such that lowest indexed project is $< k+1$, implement $(i+1)^*$ with probability $q_{(i+1)^*} - \varepsilon$ and i^* with probability $q_{i^*} + \varepsilon$;
- for all types such that lowest indexed project is $\geqslant k+1$, implement this lowest indexed project i' with probability $q_{i,i'} - \frac{\varepsilon(\alpha_{(i+1)^*} - \alpha_{i^*})}{\alpha_{i'}}$.

Now the gain for the principal is

$$\sum_{i' \leqslant i} \mu_{i'} \varepsilon(\pi_{i^*} - \pi_{(i+1)^*})$$

and the loss is

$$\sum_{i' > i} \mu_i \frac{\varepsilon(\alpha_{(i+1)^*} - \alpha_{i^*})}{\alpha_{i'}} \pi_{i'}.$$

We can see that ϵ gets canceled out, and the comparison only depends on the parameters, probabilities of the types and the payoffs.

Either the gain is greater or the loss, or they are exactly equal. If the gain is greater than the loss, then the perturbed mechanism is an IC mechanism where the principal is strictly better off, and the original mechanism cannot be optimal. If the loss is greater than the gain, then we can reverse the signs of the perturbation and achieve an IC mechanism where the principal is strictly better off again, making the previous mechanism not optimal. Finally, if the gain and the loss are exactly the same, then any perturbation would result in the same expected payoff for the principal. In this case, we can perturb the mechanism such that the expected payoff for all types is α_{k+1} and this would be an optimal mechanism as well. In addition, this optimal mechanism can be implemented in an equilibrium of the game as established in Lemma (6).

□

□