Columbia University

Department of Economics
Discussion Paper Series

Strategic Delegation By Unobservable Incentive Contracts

Levent Koçkesen
Efe A. Ok

Discussion Paper #:0102-26

Department of Economics
Columbia University
New York, NY 10027

March 2002

Columbia University
Department of Economics Discussion Paper No. 0102-26
Strategic Delegation By Unobservable Incentive Cotnracts*
Levent Koçkesen †
Efe A. Ok‡
March 2002

Abstract:

January 2002
(First version: November 1998)

Many strategic interactions in the real world take place among delegates empowered to act on behalf of others. Although there may be a multitude of reasons why delegation arises in reality, one intriguing possibility is that it yields a strategic advantage to the delegating party. In the case where only one party has the option to delegate, we analyze the possibility that strategic delegation arises as an equilibrium outcome under completely unobservable incentive contracts within the class of two-person extensive form games. We show that delegation may arise solely due to strategic reasons in quite general economic environments even under unobservable contracts. Furthermore, under some reasonable restrictions on out-of-equilibrium beliefs and actions of the outside party, strategic delegation is shown to be the only equilibrium outcome.

JEL Classification: C72, D80.

Keywords: Strategic Delegation, Unobservable Contracts, Forward Induction.

---

* We thank Kyle Bagwell, Jean-Pierre Benoit, Faruk Gul, Boyan Jovanovic, Ehud Kalai, Alessandro Lizzeri, Giuseppe Lopomo, George Mailath, Ariel Rubinstein, Andy Schotter, Rajiv Sethi, and seminar participants at Columbia, Michigan, NYU, Princeton, Rutgers, SMU, and University of Texas at Austin for helpful discussions. We also gratefully acknowledge the comprehensive comments and suggestions made by the editor Mark Armstrong and three anonymous referees, which have helped us improve the presentation substantially. The research of the first author is supported by the Alfred P. Sloan Foundation, and the second by NSF Grant No. 9808208. Both authors also acknowledge the support from the C. V. Starr Center for Applied Economics at New York University.

† Department of Economics, Columbia University, 420 West 118th Street, 1022 International Affairs Building, New York NY 10027. E-mail: lk290@columbia.edu.

‡ Department of Economics, New York University, 269 Mercer Street, New York NY 10003. E-mail: efe.ok@nyu.edu.

1 Introduction

Many strategic interactions in the real world take place among delegates empowered to act on behalf of others. Managers make strategic decisions that affect profits; sales persons have power over setting prices; and lawyers and sports agents represent their clients in bargaining processes. One of the central messages of game theory is that this could be, at least partly, due to the strategic advantage delegation may provide to the delegating party. The idea that signing binding and publicly observable contracts with a third party may serve as a beneficial commitment device goes back at least to Schelling (1960), and has been put into use in many areas of economics.[1]

However, the observability of contracts appears to be a precondition for them to play a commitment role, and for this reason, almost all applications of strategic delegation theory are couched in terms of observable contracts. The formalization of this intuition is given by Katz (1991) who showed that if contracts are unobservable, then the Nash equilibrium outcomes of a game with and without delegation coincide. In particular, delegation through unobservable contracts does not change the predicted outcome of games with a unique Nash equilibrium. These observations, in turn, call the empirical relevance of the applied delegation studies into serious question since, in most real-world transactions, the signed contracts are unobservable to the outside parties.

However, there are important strategic environments that fall outside the confines of Katz's analysis. First, Katz's model does not include the scenarios in which the outcome of the decision of delegating or not delegating is observable to the outside party. By contrast, in some models it is natural that the outside party observes the outcome of this decision simply because (s)he knows the identity of his/her opponent (even though (s)he does not know the nature of the associated contract). What is more, this may well endow the delegating party with further powers (of the forward induction type), and yield quite different insights regarding the plausibility of strategic delegation. Second, Katz's result focuses only on the alterations of the Nash equilibrium outcomes that unobserved delegation may entail. This is also quite crucial because in games where actions are taken in a sequential manner, the set of sequentially rational outcomes is generically smaller than the set of Nash equilibrium outcomes, and hence Katz's result does not tell us if these can be altered through unobserved contracts.

In fact, it is not difficult to provide examples of extensive form games in which delegation may obtain in *some* equilibrium (see Katz (1991), Fersthman and Kalai (1997)). What is not known is if there is any reason to believe that unobservable delegation has a "bite" in a large class of extensive form games. Put differently, an open problem in the literature is the determination of the conditions (on the primitives of such games) which guarantee that strategic delegation would arise in *some* equilibrium when contracts between the principals and delegates are not observable. The harder but more interesting problem is, in turn, to identify those games in which strategic delegation arises in *all* reasonable equilibria. Our objective is to extend Katz's analysis in a way that allows for

[1] A partial list includes applications in industrial organization (Vickers (1985), Fershtman and Judd (1987), Sklivas (1987), Brander and Lewis (1986)), in international trade (Brander and Spencer (1985)), in bargaining theory (Segendorff (1998)), and in monetary policy (Persson and Tabellini (1993), Jensen (1997)).

sequential games and observable delegation decisions to tackle both of these problems.

We call any finite two-person extensive form game (with perfect information) a *principals-only game* when each player (principal) plays the game himself. Let us refer to the subgame perfect equilibrium outcome of a principals-only game the *pre-delegation* outcome of that game. Given any principals-only game, we specify a (one-sided) *delegation game* as follows: in the first stage, one of the principals decides whether to play the game himself or to offer, at a cost, an incentive contract to an agent, which specifies the payoff to the agent as a function of the outcome of the game.[2] The agent, in turn, either accepts or rejects the offer. If the agent rejects the offer, the game is played between the principal and the outside party, and the delegate receives her outside option. If she accepts, then the game is played between the delegate and the outside party, and the delegate receives the payoff as specified by the contract.[3] The crucial point that distinguishes this scenario from the ones commonly considered in the literature is that, here, the outside party does *not* observe the contract offered, and knows only whether he is facing the principal or the delegate.[4]

Our main objective is to understand the nature of the sequential equilibria of this delegation game in which only one principal has the option to delegate. The first observation is that, provided that the cost of hiring an agent is relatively low, delegation may obtain in equilibrium, and this for essentially *any* principals-only game. More importantly, the outcome induced by delegation in (again essentially *any*) principals-only game can be quite different from the pre-delegation outcome of this game. This observation shows that, even under fully unobservable contracts, the act of delegation may possess commitment powers that would alter the outcome which would have obtained in the absence of delegation.

However, this finding bears an "it is possible that ..." sort of a statement, and hence provides only limited support for the presence of strategic delegation under unobservability. Yet, if we strengthen our equilibrium concept in a reasonable manner, we can understand the strategic consequences of unobserved delegation substantially better. For instance, it is possible to use a forward induction type argument to refine the equilibrium to show that delegation is essentially inevitable if the pre-delegation payoff of the delegating party is not already the best that he can obtain within a potentially large set of Nash equilibrium payoffs of the principals-only game.[5] The idea is simply that forward induction reinstates the commitment power of delegation since, under the forward induction hypothesis, the outside party interprets a delegation decision also as a signal

[2] In this paper, we study the scenarios in which only one party has the option to delegate. Understanding this simpler scenario is a prerequisite for a proper analysis of the more complicated (but obviously more realistic) case of two-sided delegation. While we will later comment briefly on how our results modify in this case, we should refer the reader to Koçkesen (1999b) for an extensive analysis of the issue.

[3] To concentrate on the strategic elements of delegation we assume that (1) the agent's sole function is to make decisions which does not require any effort, and (2) the principal and the agent are symmetrically informed.

[4] We continue to assume, however, that no renegotiation of the contract between the principal and the delegate may take place after the game commences.

[5] Clearly, it is mainly at this point that positing the realistic assumption of the observability of the identity of the delegate (and hence that of the outcome of "to delegate or not to delegate" decision of the principal) pays its dividends. This contrasts with Katz (1991) in whose model the principal does not have the option of not delegating.

about the contract that is signed. (Why would the principal pay an agent to play the game in place of him, unless he did not instruct the agent to play in a manner that improves his situation over the pre-delegation outcome even after paying the cost of hiring an agent?)

Unfortunately, forward induction type arguments that yield the above conclusion run into formal difficulties in delegation games, as we shall explain in the sequel. Consequently, we prove here the same result by using instead another intuitive equilibrium refinement, the *well-supported equilibrium*, which is based on imposing certain reasonable restrictions on the out-of-equilibrium beliefs and behavior of the players.[6] Our main result is that if there exists a Nash equilibrium outcome of the principals-only game in which (*i*) the delegating principal receives a payoff strictly greater than his pre-delegation payoff, (*ii*) the outside party behaves sequentially rationally, then in *any* well-supported equilibrium, the principal will certainly choose to delegate rather than playing the game himself, provided that the cost of hiring an agent is not too high. Moreover, this will alter the pre-delegation outcome in a way that is (strictly) beneficial for the delegating party.

The main message of the present paper may then be succinctly put as follows: To the extent that renegotiation is costly and/or limited, in a general class of economic settings, strategic aspects of delegation may play an important role in contract design, even if the contracts are completely unobservable.

The paper is organized as follows. Section 2 analyzes a simple delegation game to provide motivation for our inquiry and develop the intuition behind our main results. In Section 3, we introduce the basic nomenclature and formally introduce the equilibrium refinement that we propose here. Section 4 formally describes the economic environment within which we analyze the main question of the paper. In turn, we present our main results in Section 5, and discuss some potential extensions along with some open questions in Section 6. The proofs are contained in Section 7.

2 Motivation: A Simple Bargaining Example

In order to illustrate the basic intuition behind our results we shall first analyze a simple ultimatum bargaining game in which player 1 gives either a low offer to player 2 (denoted l) or a high offer (denoted h), and player 2 either accepts (denoted y) or rejects (denoted n) the offer. If player 1 offers l and player 2 accepts, payoffs are \$5 and \$1 for player 1 and player 2, respectively. If player 1 offers h and player 2 accepts, then player 1 receives \$1 and player 2 receives \$5. If an offer is rejected, both players receive a payoff of zero. We refer to this game as the *principals-only game*, and note that it has two Nash equilibrium outcomes (l, y) and (h, y). However, only one of these outcomes is not based on incredible threats by player 2 (i.e., it is the outcome of a subgame perfect equilibrium): player 1 offers l and player 2 accepts.

Now, assume that one of the players (principals) has the option of hiring a third player (whom we call the *agent* (or the *delegate*) and denote by A) to play the game for him. More precisely, a player can either play the game himself, that is, not hire a delegate (this action is denoted $\neg D$),

[6] Similar refinements are proposed by McLennan (1985) and Hillas (1994), and discussed extensively in Kreps (1989).

or he can offer a contract to the agent, at a cost $c > 0$, which specifies her payoffs as a function of the outcome of the principals-only game. In turn, the delegate can either accept or reject the contract. In case of rejection, player 1 and 2 play the game themselves, receive the same payoffs as in the principals-only game, except that player 2 pays the contracting cost c, and the delegate receives her outside option $\delta > 0$. If, on the other hand, she accepts the contract, then the delegate plays the game in place of the delegating player, and at any given outcome, she receives whatever the contract specifies for her, the delegating player receives the principals-only game payoff minus the cost of hiring, and the other player receives the same payoff as in the principals-only game.[7] While our description of it is not yet complete, we shall loosely refer to the resulting game as a *delegation game* in what follows.

Let us begin by observing that if it is player 1 who has the option of hiring an agent, then irrespective of which contracts are feasible and whether they are observable or not, the unique equilibrium of the delegation game would be characterized by player 1 not hiring and hence sustaining his subgame perfect equilibrium payoff. After all, \$5 is the largest possible payoff player 1 can hope for in this game, and he expects to receive this payoff if he plays the game himself.[8] Consequently, the query is interesting only when it is rather player 2 who has the option of delegating. Moreover, it is easy to see that if the cost of hiring a delegate is too high, i.e., $\delta + c > 4$, then in any sequential equilibrium of the delegation game (independent of contracts being observable or not) player 2 chooses to play the game himself. Therefore, we will analyze the case where $\delta + c < 4$.

Let us assume that there are only two contracts available to player 2, T (for tough) and W (for weak), which are specified as follows:

$$T = \begin{cases} \delta, & \text{if outcome is in } \{(l,n),(h,y)\} \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad W = \begin{cases} \delta, & \text{if outcome is in } \{(l,y),(h,y)\} \\ 0, & \text{otherwise} \end{cases}.$$

For simplicity, we will assume that the delegate accepts any contract that yields her at least her outside option as expected payoff. (This assumption can be relaxed, provided that one enlarges the contract space suitably.) Therefore, the delegate accepts any contract offer in the menu $\{T, W\}$ since, irrespective of player 1's offer, there is always an action which would earn her δ.

If the contract signed between player 2 and his delegate were observable to player 1, then the unique subgame perfect equilibrium outcome of the delegation game would have player 2 offering the contract T, player 1 offering h and the delegate accepting the offer. This is of course nothing but a simple demonstration of the beneficial commitment effects of *observable* delegation. Things get a bit more complicated, however, if we (realistically) assume that only the decision to hire a delegate or not is observable by player 1, not the contract offered. The description of the game

[7] Of course, the delegating player could himself earn an outside option by delegating. We, however, assume that the outside option of this player is zero to make the analysis interesting. Clearly, if this outside option was large relative to the potential payoffs in the game, delegation would obtain due to nonstrategic reasons.

[8] This observation is true only in the case of one-sided delegation. If the delegation game is two-sided, then, as Koçkesen (1999b) shows, there is an equilibrium in which both players randomize between delegating and not delegating.

becomes complete under this informational assumption; we refer to this game as the *delegation game* and depict its basic structure in Figure 1.[9]

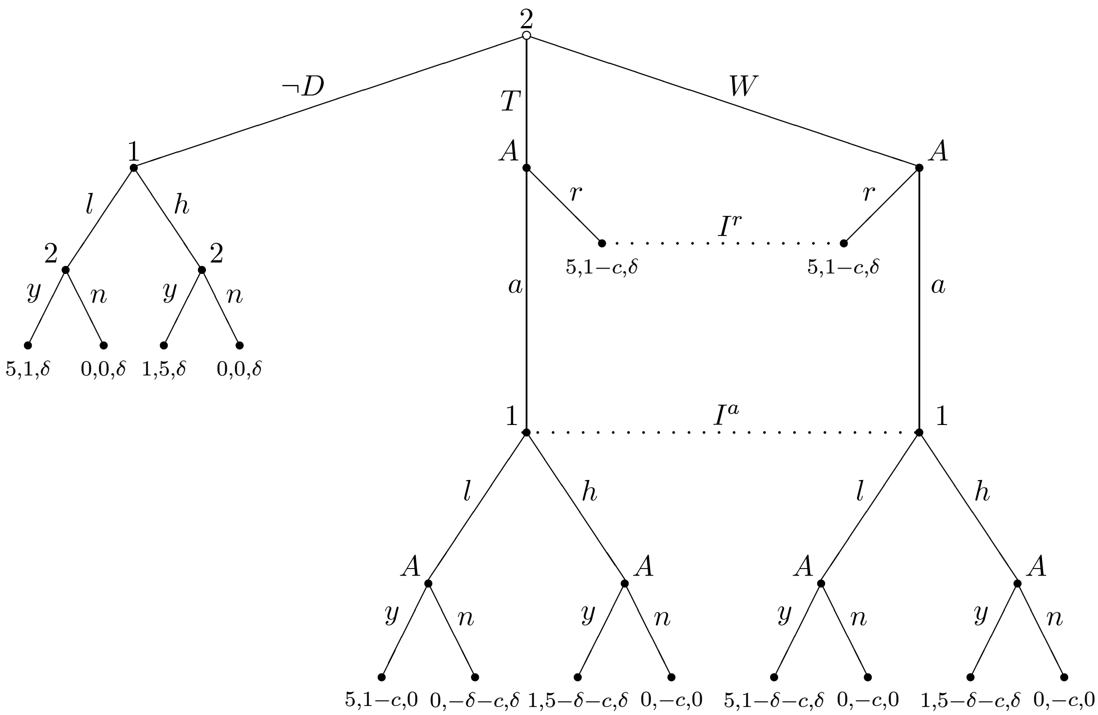

Figure 1: Delegation in Ultimatum Bargaining.

There are two types of sequential equilibria of this game. The first type is characterized by player 2 not delegating, and the second is characterized by player 2 choosing an action other than $\neg D$. In all equilibria of the first type, following player 2's action, player 1 offers l and player 2 accepts. At the out-of-equilibrium information set following an accepted contract (denoted I^a in Figure 1), player 1 believes that contract W has been offered with at least probability $1/5$ and he plays l with at least probability $1-(\delta+c)/4$. In the second type of equilibrium, player 2 places at least probability $4/5$ on contract T and, at the information set I^a, player 1 plays h.

While all of these equilibria are in fact trembling-hand perfect (Katz (1991), Fersthman and Kalai (1997)), it is still possible to take issue with the plausibility of a first type of equilibrium. To simplify the discussion, let us first allow only for pure strategies. (The case of mixed strategies are taken up in Subsection 3.2; see Example 2.) The point is that, in this case, we can apply a natural forward induction argument to "kill" any of the first type of equilibria. In particular, no such equilibrium survives the forward induction test proposed by van Damme (1989). In any

[9] For simplicity, we truncated the branches of the game tree following the delegate's action r and replaced them by the equilibrium outcomes which would be realized if the play were ever to reach there.

pure strategy equilibrium that passes this test, player 2 offers the contract T and player 1 plays h: forward induction ensures delegation even if the contracts are unobservable.

Another observation which points out to what is unreasonable about the first type of equilibrium is that the contract W, which aligns the incentives of player 2 and the agent, is not offered in *any* pure strategy equilibria of the game since player 2 would be better off by playing the game himself rather than hiring an agent through the contract W. Yet, the only way one can support an equilibrium in which player 2 plays $\neg D$ is by assuming that player 1 believes at I^a that the agent he is facing has been offered nothing but the contract W!

In the final analysis, whether such beliefs are reasonable or not are formally captured in the equilibrium concept one adopts to "solve" the game. The out-of-equilibrium beliefs supporting an equilibrium of the first type (in which player 2 plays $\neg D$) are justified in a sequential equilibrium simply because player 1 thinks that player 2 has made a mistake, without trying to make further inferences regarding player 2's possible play which caused the information set I^a to be reached. Suppose, in contrast, that player 1 rather reasons, upon facing a delegate unexpectedly, that it is actually he who made a mistake in assuming that a first type of equilibrium is accepted as the current norm. (Kreps (1989) calls refinements based on this line of reasoning "mistaken theory" refinements; see Section 3.) He may then well conclude that player 2 is playing according to some other equilibrium in which a delegate is hired. But, in no such equilibrium player 2 offers the contract W, and therefore, so player 1 reasons, W cannot be the contract that is signed. Given that his beliefs put probability zero on contract W, player 1's optimal action is h and hence player 2 strictly prefers to delegate in any equilibrium which survives a "mistaken theory" refinement.

It is not clear which interpretation ("mistakes" or "mistaken theories") is more plausible in general. The answer is likely to depend on the situation being analyzed. In delegation games, however, there is reason to believe that the second interpretation is more convincing. These games depict situations in which individuals decide whether to hire someone to act on their behalf or not in a strategic interaction. It would not be reasonable to think that such a decision, which in reality requires time and effort, takes place without careful deliberation. Hiring someone is a costly action and it is unlikely to take place as a result of sheer irrationality or a simple mistake. Arguably, therefore, in all "reasonable" equilibria of the above delegation game, player 2 hires a delegate and offers a contract so that the equilibrium outcome is different from the subgame perfect equilibrium outcome of the principals-only game. Hence, we contend that delegation is likely to ensue even by means of *unobservable* contracts; this is the main thesis we shall defend formally in this paper.

In the rest of the analysis, we shall consider arbitrary extensive form principals-only games with perfect information, and derive necessary and sufficient conditions for the existence of commitment value of delegation under unobservable contracts. More precisely, we will show that if there exists a Nash equilibrium outcome of the principals-only game in which (i) the delegating principal receives a payoff strictly greater than her subgame perfect equilibrium payoff, and (ii) the outside party behaves sequentially rationally, then in any well-supported equilibrium, the principal will certainly choose to delegate rather than playing the game himself, provided that the cost of hiring an agent is low enough.

3 Preliminaries

3.1 Basic Nomenclature

Following Osborne and Rubinstein (1994), we define a finite-horizon *extensive form game* as a collection

$$\Upsilon \equiv [N, H, P, (\mathcal{I}_i, \pi_i)_{i \in N}].$$

Here N denotes a finite *set of players*, and H stands for a finite comprehensive set of finite sequences interpreted as the *set of all histories*.[10] An history h is said to be *terminal* if $(h, a) \notin H$ for any $a \neq \emptyset$; we denote by Z the set of all terminal histories. The function $\pi_i : Z \to \mathbb{R}$ is the *payoff function* of player i, and the function $P : H \backslash Z \to N$ is the *player function*. If $P(h) = i$, we understand that i moves immediately after history h and chooses an action from the set $A(h) \equiv \{a \neq \emptyset : (h, a) \in H\}$. For each i, $\mathcal{I}_i$ is a partition of $H(i) \equiv \{h \in H : P(h) = i\}$ such that $A(h) = A(h')$ whenever $h, h' \in I \in \mathcal{I}_i$. Consequently, without ambiguity, we may write $A(I)$ ($P(I)$, resp.) instead of $A(h)$ ($P(h)$, resp.) for any $h \in I$. Any member of $\mathcal{I}_i$ is called an *information set* for player i. If all information sets in Υ are singletons, we say that this game is with *perfect information*, and omit information partitions in its definition. The subgames of Υ are defined in the usual way.

A *behavioral strategy* for player i is defined as a set of independent probability measures $\beta_i \equiv \{\beta_i[I] : I \in \mathcal{I}_i\}$ where $\beta_i[I]$ is defined on $A(I)$. One may write $\beta_i[h]$ for $\beta_i[I]$ for any $h \in I$ with the understanding that for any h and h' that belong to the same information set, we have $\beta_i[h] = \beta_i[h']$. The set of all behavioral strategies of player i is denoted $S_i(\Upsilon)$, whereas $S(\Upsilon) \equiv \times_{i \in N} S_i(\Upsilon)$. We denote the set of all *Nash* and *subgame perfect* equilibria of Υ in behavioral strategies by $NE(\Upsilon)$ and $SPE(\Upsilon)$, respectively.

By a *system of beliefs*, we mean a set $\mu \equiv \{\mu[I] : I \in \mathcal{I}_i \text{ for some } i\}$, where $\mu[I]$ is a probability measure on I. We denote the set of all systems of beliefs by $B(\Upsilon)$. A 2-tuple $(\beta, \mu) \in S(\Upsilon) \times B(\Upsilon)$ is called an *assessment*. An assessment is said to be a *sequential equilibrium* if it satisfies the properties of *consistency* and *sequential rationality* (see Section 7.A for formal definitions). We denote the set of all such assessments as $SE(\Upsilon)$. The set of all equilibria in $SE(\Upsilon)$ that reach to the information set I with positive probability is denoted $SE(\Upsilon; I)$.

3.2 Well-Supported Bayesian Equilibria

We next introduce a refinement of sequential equilibria that will play a central role in this paper. Like many others, this refinement too is based on imposing certain restrictions on the out-of-equilibrium beliefs and strategies of the players. Informally put, it leads us to those "well-supported" sequential equilibria that envisage that at each out-of-equilibrium information set, the beliefs and behavior of the players are consistent with at least one sequential equilibrium that admits this information set on its equilibrium path, provided that such an equilibrium exists.

[10] By *comprehensiveness* of H, we mean that $\emptyset \in H$, and, for any integer $k \geqslant 1$, $(a^1, ..., a^k) \in H$ whenever $(a^1, ..., a^{k+1}) \in H$. The *length* of a history $h = (a^1, ..., a^k)$ is defined to be k and denoted by $|h|$. As a convention, we take $|\emptyset| = 0$ and let $(h, \emptyset) = h$ for any $h \in H$.

Here we shall give the formal definition of our refinement as it applies only to the class of all extensive form games which do not possess an imperfect information subgame which is distinct from the original game. We denote this class of games by $\mathcal{G}$. Our formal treatment is distilled to its simplest form in the context of such games, thereby making the intuition behind the refinement proposed here transparent. Moreover, the set of games which is the focus of this paper is a subset of $\mathcal{G}$, so this simplified treatment does not cause a loss of generality for our purposes.[11]

For any $\Upsilon \in \mathcal{G}$ and any $\beta \in S(\Upsilon)$, we define $\mathcal{I}_i(\beta)$ as the set of all information sets of player i that are reached by β with positive probability, and for any information set I, we let $\mathcal{I}_i(I)$ be the set of all information sets of i that follow I.[12] On the other hand, $\mathcal{J}(\beta)$ stands for the set of all nonsingleton information sets that could be reached with the shortest sequence of actions after a deviation from $\beta_{P(\emptyset)}[\emptyset]$, while they are surely not reached when $\beta_{P(\emptyset)}[\emptyset]$ is played.[13]

Definition 1. Let $\Upsilon \in \mathcal{G}$. A sequential equilibrium $(\beta, \mu) \in SE(\Upsilon)$ is said to be **well-supported** if, and only if, for each $I \in \mathcal{J}(\beta)$, either $SE(\Upsilon; I) = \varnothing$ or there exists a $(\beta', \mu') \in SE(\Upsilon; I)$ such that

$$\mu[J] = \mu'[J] \quad \text{and} \quad \beta_{P(I)}[J] = \beta'_{P(I)}[J] \tag{1}$$

for all $J \in \mathcal{I}_{P(I)}(I)$. We denote the set of all well-supported sequential equilibria of Υ by $SE_{\text{w-s}}(\Upsilon)$.

To clarify things, let us take a two-player game $\Upsilon \in \mathcal{G}$. Suppose that player 1 moves first in this game and suppose that I is a first nonsingleton information set of player 2 which is on the out-of-equilibrium path (i.e., $I \in \mathcal{J}(\beta)$). Player 1's move may indeed be a part of a sequential equilibrium (β, μ), provided that it is suitably supported by beliefs and the continuation strategy of player 2 at the out-of-equilibrium information set I. In a sequential equilibrium, if I is ever reached, player 2 could interpret this as a simple "mistake." This, however, goes counter to the idea that player 1's deviation could be evaluated by player 2 as containing information about player 1's past (unobserved) actions.

How will player 2 reason when he finds himself at I which was not supposed to be reached in the equilibrium that is being played? She may plausibly think that player 1 is after coordinating on a different equilibrium, provided that an equilibrium that reaches I exists (i.e., $SE(\Upsilon; I) \neq \emptyset$). Of course, if $SE(\Upsilon; I) = \emptyset$, then there is no such plausible explanation of 1's deviation, and hence no restriction is imposed on 2's beliefs and behavior at I. On the other hand, if there exists exactly one such equilibrium, then (β, μ) is well-supported only if the beliefs and the strategy of player 2 following this information set accord with what is specified by (β, μ). This is precisely the requirement embodied in (1). Alternatively, if there exist more than one such equilibrium, then

[11] It is not difficult to generalize the definition of well-supported equilibrium to the class of all finite extensive form games by means of an inductive argument that admits the definition we present here as the first step of the induction process. (See Koçkesen (1999a), pp. 91-94.)

[12] Formally speaking, $\mathcal{I}_i(I) \equiv \{J \in \mathcal{I}_i : h'' \in J \text{ iff } h'' = (h, h') \text{ for some } h \in I\}$. Notice that $I \in \mathcal{I}_i(I)$.

[13] To define $\mathcal{J}(\beta)$ formally, define $S(\beta_{P(\emptyset)}[\emptyset])$ as the set of all behavioral strategy profiles in which the first mover in the game plays $\beta_{P(\emptyset)}[\emptyset]$ at the initial node. Define next $\mathcal{Q}(\beta_{P(\emptyset)}[\emptyset])$ as the set of all non-singleton information sets that do not belong to $\cup_{i \in N} \cup_{\beta' \in S(\beta_{P(\emptyset)}[\emptyset])} \mathcal{I}_i(\beta')$. We have $I \in \mathcal{J}(\beta)$ if and only if $I \in \mathcal{Q}(\beta_{P(\emptyset)}[\emptyset])$ and for any $J \in \mathcal{Q}(\beta_{P(\emptyset)}[\emptyset])$ there do not exist h' and $h'' \neq \emptyset$ such that $h' \in J$ and $(h', h'') \in I$.

(β, μ) is well-supported only if the continuation beliefs and the strategy of player 2 agrees with that of at least one such equilibrium.[14]

Example 1. (Battle of the sexes with an outside option)[15] Consider the game $\Upsilon \in \mathcal{G}$ depicted in Figure 2. In this game there are two types of equilibria; one in which $\beta_1[\emptyset](O) = 1$ and one in which $\beta_1[\emptyset](T) = 1$. For the first type of equilibria, we must have either $\beta_2[I](R) = 1$ and $\mu[I](T) < 3/4$ or $\beta_2[I](L) \leqslant 2/3$ and $\mu[I](T) = 3/4$. The unique second type of equilibrium, however, has it that $\beta_2[I](L) = 1$ and $\mu[I](T) = 1$. For the first type of equilibria we have $\mathcal{J}(\beta) = \{I\}$, that is, I is the only out-of-equilibrium information set to check. Since only the second type of equilibrium reaches I, if a first type of equilibrium is well-supported, it must have $\beta_2[I](L) = 1$ and $\mu[I](T) = 1$, that is, part (c) of the definition must hold. Since this is not the case, we conclude that none of the first type of equilibria is well-supported. The unique well-supported equilibrium outcome of Υ is then (T, L) with the payoff profile $(3, 1)$. Notably, this is also the unique outcome that passes the forward induction test of van Damme (1989).[16] $\|$

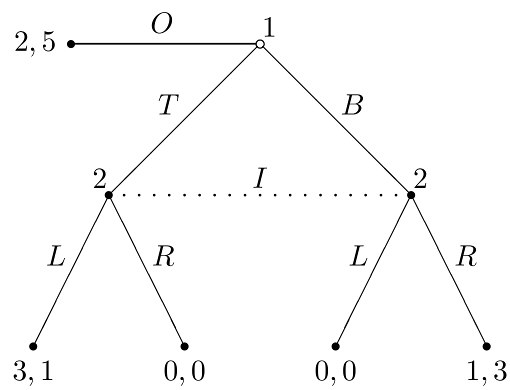

Figure 2: Battle of the Sexes with an Outside Option

Example 2. (A simple bargaining game (cntd.)**)** Consider the simple bargaining game discussed in Section 2. It is easy to show that any equilibrium in this game in which player 2 chooses $\neg D$ with positive probability (first type of equilibria) is not well-supported. In any such equilibrium, we have $\beta_2[\emptyset](\neg D) > 0$, $\mathcal{J}(\beta) = \{I^a\}$ and $\beta_1[I^a](l) > 0$, whereas in all equilibria

[14] In games with more than two players, there is reason to view the well-supportedness requirement as too stringent, for in such games it may be desirable to restrict the beliefs/behavior of players at an off-equilibrium information set only when *all* players are aware that this set lies on an off-the-equilibrium path. While this sort of a weakening is easy to formalize, because such an issue does not arise within the general class of delegation games that we shall be concerned with here (there is only one party who needs to identify the off-equilibrium nonsingleton information sets in such games), we choose to adopt the simpler formulation given in Definition 1. Koçkesen and Ok (1999) define and discuss the weaker version of the well-supported equilibrium concept.

[15] This example is due to Kohlberg and is probably the most commonly used game in motivating the basic idea behind the notion of forward induction. It is also experimentally analyzed by Cooper et al. (1993) who provided mixed evidence in support of the forward induction hypothesis.

[16] The last two observations remain true if we break up the moves of player 1 into two component parts so that he first decides whether to choose O or N, and only after he chooses N he decides between T and B.

that reach information set I^a (the second type of equilibria) player 1 offers l with zero probability. On the other hand, all the equilibria of the second type, i.e., delegation equilibria, are trivially well-supported.

It can be checked that the latter equilibria do not survive the iterated elimination of weakly dominated strategies, and hence do not form a strategically stable set. Yet, if one slightly perturbs the game so that there is a positive probability, however small, that the offered contracts are observed, one can show that delegation equilibrium survives the iterated elimination of weakly dominated strategies. ‖

We next comment briefly on how the well-supportedness criterion relates to some other major equilibrium refinements proposed in the literature.

Remarks. (1) (*Well-Supportedness, Forward Induction, and the "Mistaken Theory" Refinements*) As noted earlier, Kreps (1989) refers to solution concepts that are based on the idea that deviations should, when possible, be viewed as players coordinating on other equilibria as "mistaken theory" refinements. A version of this approach was first developed by McLennan (1985) who argued that "deviations from the equilibrium path are more probable if they can be explained in terms of some confusion over which sequential equilibrium is "in effect"." (McLennan (1985), p. 891). This is also the leading motivation behind the *forward-induction* refinements of van Damme (1989) and Al-Najjar (1995), as well as the refinements proposed by Hillas (1994) and our well-supportedness concept.[17]

The forward induction refinement proposed by van Damme (1989) may be found too weak due to its reliance on the *viability* requirement. (This condition postulates that a deviation is viewed as an unambiguous signal *only if* there is a unique continuation equilibrium that makes the deviation profitable for the deviating party.) For this reason, Al-Najjar (1995) has proposed a stronger refinement that does away with this requirement. Unfortunately, as shown by examples in that paper, this refinement is in a sense "too strong" for it may well eliminate certain reasonable equilibria. Intuitively speaking, the well-supported equilibrium concept falls in between these two refinements, at least in the present setup. In particular, van Damme's forward induction criterion is not sufficient for our second main result (even if we restrict the analysis to pure strategies). On the other hand, Al-Najjar's refinement delivers that result easily, albeit at the cost of potentially eliminating some equilibria (in which delegation does not take place) in an ad hoc manner.[18] By contrast, while well-supported equilibrium is weaker than Al-Najjar's refinement,[19] it is strong

[17] As a caveat, however, we note that some other conceptualizations of forward induction (that are linked inherently to the notion of the iterated elimination of weakly dominated strategies), such as that of Kohlberg (1990), is not a "mistaken theory" refinement.

[18] Observe that, in contrast to the pure strategy analysis given in Section 2, van Damme's forward induction refinement does not yield the result reported in Example 2 (since there are more than one sequential equilibrium with delegation that leave to player 2 a payoff strictly higher than his subgame perfect equilibrium payoff). Al-Najjar's refinement, on the other hand, agrees in this game fully with the well-supported equilibrium.

[19] This claim is formal in the context of delegation games that we consider here. One can also easily check that our refinement behaves well in the examples that Al-Najjar has considered to illustrate the shortcomings of his refinement.

enough to eliminate all "no-delegation equilibria."

Perhaps the best formalization of the "mistaken theory" approach is provided by Hillas (1994). The well-supported equilibrium is quite close to the refinements considered by Hillas both in motivation and in formalization. For the record, we report that Hillas's first two refinements (where restrictions are imposed either only on beliefs or only on choice behavior at every information set) are not strong enough for our purposes. His third refinement puts restrictions on *both* beliefs *and* behavior, and is strong enough for our results to go through. The well-supported equilibrium concept is, however, not only weaker than this refinement, but also much simpler to apply within the delegation games that we study here.

(2) (*Well-Supportedness vs. Strategic Stability*) The main example analyzed by van Damme (1989, pp. 485-87) shows that strategically stable equilibria (Kohlberg and Mertens (1986)) need not satisfy his forward induction criterion. Since the requirement of well-supportedness chooses precisely the equilibrium chosen by van Damme's forward induction concept in his example, we may also conclude that strategic stability does not imply well-supportedness. Conversely, a well-supported equilibrium may not survive iterated elimination of weakly dominated strategies, and hence need not be strategically stable. (See Example 2.) This discrepancy between strategic stability and well-supportedness is not surprising. Like most other similar "mistaken theory" refinements, well-supportedness is logically independent of the notion of the iterated elimination of weakly dominated strategies, whereas strategic stability admits this notion as a prerequisite. ‖

We think of well-supportedness as a reasonable (and a somewhat weak) refinement of the sequential equilibria. Rather than arguing for the superiority of our refinement over others, however, we subscribe here to the view that "the validity of a particular refinement for the analysis of a particular economic issue may depend on the setting of that issue in ways that go beyond the formal game-theoretic model that is adopted" (Kreps (1989), p. 7.) The main objective of this paper is to use the concept of "well-supportedness" in an economic setting in which it is particularly sensible and only mildly demanding. Moreover, in this context, we shall see that it allows one to obtain considerable insight with regard to the underlying economic problem. We turn next to describing this problem in detail.

4 One-Sided Delegation Environments

In this section we shall introduce a general environment in which we shall study the problem of delegation by unobservable incentive contracts. As one might expect, the framework we outline below admits the simple bargaining-delegation model studied in Section 2 as a special case.

We begin by fixing an arbitrary finite perfect information *principals-only* game

$$\Gamma = [\{1,2\}, H, P, (\pi_1, \pi_2)].$$

We assume that this game has a unique subgame perfect equilibrium outcome, which we consider as the *pre-delegation* outcome of the environment prior to delegation. We denote the (pre-delegation) expected payoff of player i in equilibrium by Π_i^{SPE}. The set of all Nash equilibrium payoffs of i is in turn denoted by $\mathbf{\Pi}_i^{\mathrm{NE}}(\Gamma)$. In what follows, we assume that $\mathbf{\Pi}_i^{\mathrm{NE}}(\Gamma)$ is a finite set for each $i = 1, 2$.[20]

Let $NE_i^*(\Gamma)$ denote the set of Nash equilibria of Γ in which the behavioral strategy of player i is sequentially rational after *every* history.[21] We denote the set of all expected payoffs for j that correspond to the strategy profiles in $NE_i^*(\Gamma)$ by $\mathbf{\Pi}_j^{\mathrm{NE}_i^*}(\Gamma)$.

Suppose that player 2 is contemplating about hiring an agent to play the game Γ in place of him. The *outside option* of this agent, henceforth called player A, is a constant $\delta > 0$. That is, player A receives δ dollars with certainty if she rejects the contract offered by player 2. Moreover, we assume that player 2 incurs a contracting cost of $c > 0$ dollars in case he decides to delegate.

In the literature on strategic contract design, a contract is sometimes defined as a mapping that specifies a level of payment for each *strategy* of the agent in $S_2(\Gamma)$. Such a contract provides an extensive amount of control to the principal, and usually simplifies the analysis considerably. However, especially when randomization is allowed, this definition would lead one to view a contract as an unrealistically complicated object. Moreover, it is not at all clear how a principal could in general "observe" the randomized strategy choice of an agent, which, to be able to submit the agent's compensation, he must. At the very least, this would necessitate to extend the model to account for private monitoring of the agent.

In our model a *contract* is a function that maps the finite set of terminal nodes Z of the game Γ to the set of payments. In fact, it will be sufficient here to focus on those simple contracts that pay the agent either her outside option or nothing at all. Thus, the *contract space* in our model will be designated as $\{0, \delta\}^Z$. Evidently, any member of $\{0, \delta\}^Z$ is an *incentive contract* that can be conditioned only on the pure outcomes of the game rather than the delegate's strategy.[22] While working with such contracts introduces a number difficulties regarding the formal analysis, it brings us a step closer to realism and avoids worrying about issues related to the "monitoring" of the agents since the pure outcomes of the game Γ are observable.

The primitives of our model is thus the game Γ, the outside option $\delta > 0$ and the contracting cost $c > 0$. We thus refer to the 3-tuple $[\Gamma, \delta, c]$ as a *one-sided delegation environment.* Such an environment naturally induces a delegation game

$$\Lambda(\Gamma, \delta, c) \equiv [\{1, 2, A\}, H^*, P^*, (\mathcal{I}_i^*, \pi_i^*)_{i \in \{1,2,A\}}]$$

which is a 3-person extensive form game that will on occasion be simply denoted as Λ. The game begins with player 2 deciding between taking the action of not delegating (denoted $\neg D$) and an action of attempting to delegate by offering a contract $f \in \{0, \delta\}^Z$, which the agent A may accept

[20] Finiteness of $\mathbf{\Pi}_i^{\mathrm{NE}}(\Gamma)$ is an assumption that helps us avoid some technical redundancies and is generically true for finite extensive form games (Kreps and Wilson (1982)).

[21] Formally, $\beta^* \in NE_i^*(\Gamma)$ iff $\beta^* \in NE(\Gamma)$ and, for each $h \in H(i)$, we have $\beta_i^* \in \arg\max_{\beta_i \in S_i(\Gamma)} \Pi_i(\beta_i, \beta_{-i}^* \mid h)$ where $\Pi_i(\cdot \mid h)$ is the expected payoff of player i conditional on history h being reached.

[22] The distinction is reminiscent of the distinction made by Fershtman and Kalai (1997) between *incentive* and *instructive* delegation. (In the former case, the delegate is given an incentive scheme correlating her payments with performance, and in the latter case, she is directly given a strategy that she must follow).

(denoted a) or reject (denoted r).[23] If player 2 chooses not to delegate, or he chooses to offer a contract but this contract is rejected by A, then Γ is played between players 1 and 2. But if player 2 offers a contract f that agent A accepts, then, A (instead of player 2) plays the game Γ against player 1. Therefore, there are three types of histories in H^*. Histories that pertain to the no delegation decision (e.g. $(\neg D, h)$) can be identified with those of the principals-only game (that is, with h). Similarly, histories that pertain to a rejected contract offer (e.g. (f, r, h)) can be identified with those of Γ. On the other hand, on the path of a history like (f, a, h), the contract offer f is accepted, and the game reached to $h \in H$ with agent A playing in place of player 2.[24]

Of course, with or without delegation, the play in Λ must lead to an outcome of the principals-only game. Indeed, to any terminal history $z^* \in Z^*$, there corresponds a unique outcome $z \in Z$ in Γ such that $z^* \in \{(\neg D, z), (f, r, z), (f, a, z)\}$. We shall refer to this terminal history z as a *pure outcome induced by* z^* in Γ.

It is crucial to recognize that while player 1 observes whether or not a contract is accepted, that is, he always knows the *identity* of her opponent, he does not observe which contract is accepted (or rejected). Hence, once a contract is accepted, player 1 does not know the payoff function of his opponent (i.e., of player A). Players 2 and A, on the other hand, possess perfect information throughout the game so that all of their information sets are singletons.[25]

Next we need to specify the payoff functions of the players. Since player 1 is not involved with any sort of a delegation activity, we have $\pi_1^*(\neg D, z) \equiv \pi_1^*(f, \theta, z) \equiv \pi_1(z)$ for all $z \in Z$, $\theta \in \{a, r\}$, and $f \in \{0, \delta\}^Z$. Similarly, the payoffs of player 2 would not be altered if he chooses not to delegate, that is, $\pi_2^*(\neg D, z) \equiv \pi_2(z)$ for all $z \in Z$. On the other hand, player 2 incurs the cost c if he chooses to offer a contract, and pays the promised compensation to the agent in case a contract is signed. Therefore,

$$\pi_2^*(f, \theta, z) \equiv \begin{cases} \pi_2(z) - f(z) - c, & \text{if } \theta = a \\ \pi_2(z) - c, & \text{if } \theta = r \end{cases}$$

for all z and f. Finally, the delegate's payoffs are determined as $\pi_A^*(\neg D, z) \equiv \pi_A^*(f, r, z) \equiv \delta$ and $\pi_A^*(f, a, z) \equiv f(z)$ for all z and f. This completes the description of the delegation game $\Lambda(\Gamma, \delta, c)$.

In what follows, we shall investigate the sequential equilibria of Λ. However, to avoid certain technical difficulties, we shall restrict ourselves to a particular subclass of $SE(\Lambda)$ in which the equilibrium strategy of the agent A is to accept a contract whenever the expected value of the contract equals his outside option δ. So, henceforth, all references to a sequential equilibrium of

[23] An alternative model would have player 2 choosing first between not delegating and delegating, and then choosing a contract if he decides to delegate. All our results go through under this alternative modeling assumption.

[24] For concreteness, we note that $H^* \equiv (\{\neg D\} \times H) \cup (\{0, \delta\}^Z \times \{a, r\} \times H)$. The player function P^* is defined on the nonterminal histories in H^* by letting $P^*(\varnothing) \equiv 2$, $P^*(f) \equiv A$, $P^*(\neg D, h) \equiv P^*(f, r, h) \equiv P(h)$, and finally, $P^*(f, a, h) \equiv A$ if $P(h) = 2$, and $P^*(f, a, h) \equiv 1$ if $P(h) = 1$.

[25] Thus, the information partition of player 1 is

$$\mathcal{I}_1^* \equiv \{\{(\neg D, h)\} : h \in H(1)\} \cup \{\cup_f \{(f, \theta, h)\} : \theta \in \{a, r\},\ h \in H(1)\},$$

whereas $\mathcal{I}_2^* \equiv \{\emptyset\} \cup \{\{(\neg D, h)\} : h \in H(1)\} \cup \{\{(f, r, h)\} : f \in \{0, \delta\}^Z, h \in H(2)\}$ and $\mathcal{I}_A^* \equiv \{\{(f, a, h)\} : f \in \{0, \delta\}^Z, h \in H(2)\}$.

Λ should be taken to apply only to this subclass. While this sort of a tie-breaking assumption is commonly invoked in the related literature, it nevertheless amounts to a somewhat arbitrary equilibrium refinement, and hence may justly be found objectionable. Fortunately, all of our results remain valid *verbatim* when this restriction is dispensed with, provided that one allows for an infinite contract space (like $\mathbb{R}^Z$), and define strategies and beliefs in terms of simple probability measures. However, the analysis of the resulting model is substantially more complicated than the present one while it does not provide new insights. For this reason, we adopt here the standard tie-breaking postulate stated above without apology, and refer the reader to Koçkesen and Ok (1999) for the more general analysis.

5 The Main Results

Fix a one-sided delegation environment $[\Gamma, \delta, c]$. As noted earlier, our main objective here is to understand the nature of the equilibria of the induced delegation game $\Lambda(\Gamma, \delta, c)$ as it pertains to the implications of the possibility of delegation. Thus, the first question we need to address is if delegation takes place in equilibrium at all, while the second question is if the *pre-delegation* equilibrium *outcome* of the game Γ is altered, provided that at least some degree of delegation takes place in equilibrium. The literature on delegation since the influential contribution of Katz (1991) exhibits clearly the contention that neither of these questions have an affirmative answer. Indeed, the analysis of Katz (1991) culminates in showing that the Nash equilibrium outcomes of the principals-only game, and only these outcomes, can be reached via unobserved delegation in the Nash equilibria of the delegation game which satisfies a weak sequential rationality constraint. Consequently, a sequentially rational equilibrium outcome of the delegation game has to be a Nash equilibrium outcome of the principals-only game. More precisely, in the present setting, we have

Proposition 1. *If the pure outcome z^* is reached with positive probability in a sequential equilibrium of the delegation game $\Lambda(\Gamma, \delta, c)$, then the pure outcome induced by z^* in the principals-only game Γ can be reached with positive probability in a Nash equilibrium of Γ in which player 1 plays sequentially rationally at every history.*

The intuition behind this result is quite simple. Since the contract offered by player 2 is not observable, the outside party, i.e., player 1, cannot condition his strategy on the contract. Therefore, rationality of player 1 implies that he must be offering a contract that induces the agent to best respond to player 1 in terms of player 2's preferences. Similarly, rationality and consistency imply that player 1 must best respond to the agent's strategy that is induced by such an optimal contract. This entails that any sequential equilibrium outcome of the delegation game must be a Nash equilibrium outcome of the principals-only game. The last part of the claim which states that player 1's strategy must be sequentially rational at every history is just an easy implication of sequential rationality.

Moreover, it can be shown that the set of expected payoffs of the principal 2 (gross of the cost of hiring) obtained in a sequential equilibrium of $\Lambda(\Gamma, \delta, c)$ lies in the convex hull of $NE_1^*(\Gamma)$ payoffs of principal 2 which are at least as large as his subgame perfect equilibrium payoff in game Γ. Thus, the possibility of delegation does not alter in a payoff-relevant way the set of *Nash equilibrium outcomes* of the principals-only game. In particular, we have the following negative result.

Corollary 1. *If Γ is a simultaneous move game with a unique Nash equilibrium, then the outcome of this equilibrium is identical to that of any sequential equilibrium of the delegation game $\Lambda(\Gamma, \delta, c)$.*[26]

It is undeniable that Corollary 1 (the main thrust of which should be credited to Katz (1991)) creates severe difficulties for the well-known (perfect information) delegation results that concern simultaneous move principals-only games such as those of Vickers (1985), Fersthman and Judd (1987) and Sklivas (1987). However, as noted by Fersthman and Kalai (1997), its implications become limited when we shift our focus to games with sequential moves.

We show below that there is a formal sense in which Proposition 1 and Corollary 1 are incomplete descriptions of matters in the case of extensive form principals-only games. While, if the cost of delegation $\delta + c$ is too high, delegation does not obtain for such games either, for low enough $\delta + c$, the situation is vastly different. In particular, it turns out that, for small $\delta + c$, any Nash equilibrium payoff of the delegating principal, which is at least as large as his subgame perfect equilibrium payoff, and which can be obtained via a sequentially rational strategy of the outside party can also be obtained as a sequential equilibrium payoff of the delegation game. This is, in fact, a complete characterization of the set of all sequential equilibrium payoffs of $\Lambda(\Gamma, \delta, c)$, and is our first main result.

Theorem 1. *There exists an $\ell > 0$ such that $\delta + c < \ell$ implies that*

$$\mathbf{\Pi}_2^{\mathrm{SE}}(\Lambda(\Gamma, \delta, c)) = \{\Pi_2 \in \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\mathrm{SPE}}\},$$

where $\mathbf{\Pi}_2^{\mathrm{SE}}(\Lambda(\Gamma, \delta, c))$ is the set of all sequential equilibrium (expected) payoffs for player 2.

This result says that delegation *may* alter the equilibrium outcome of an extensive form game that would obtain in the absence of delegation in a very large class of games when the cost of delegation is small. Notice that unobserved delegation expands the set of equilibrium outcomes to include those in which the delegate best responds to the outside party's strategy from the perspective of the principal's preferences, but not necessarily in a sequentially rational manner, whereas the outside party plays sequentially rationally. In that sense delegation achieves a commitment effect even under unobservable contracts by freeing the principal from the straitjacket of sequential rationality, which in turn enables him to issue threats (that would otherwise be incredible) via his delegate. A more detailed intuition behind this theorem is provided at the end of this section.

The next question is if one can strengthen the argument for delegation by showing that reasonable requirements of rationality ensure that strategic delegation is bound to alter the equilibrium

[26] Formally speaking, Corollary 1 is not implied by Proposition 1 since in the latter result we have assumed that Γ is an extensive game with perfect information. However, as we shall stress in Section 6 again, Proposition 1 also applies to any extensive game with a unique sequential equilibrium (and hence to normal form games with a unique equilibrium). The proof of this version of Proposition 1 is analogous to the one we shall provide in Section 7.

outcome of an extensive form principals-only game in a way that benefits the delegating principal. Put more precisely, the query at hand is this: *Given a one-sided delegation environment* $[\Gamma, \delta, c]$, *is there any well-supported equilibrium of* $\Lambda(\Gamma, \delta, c)$ *in which delegation does not take place?*

Of course, the requirement of well-supportedness is indeed needed in the statement of this query, for we know that there exist sequential equilibria in which delegation does not obtain. (Recall Section 2.) As noted earlier, a major point of the present paper is that such equilibria are unreasonable, and a suitable forward induction and/or out-of-equilibrium behavior restriction argument will eliminate all equilibria that envisage a neutral role for delegation. We propose the notion of "well-supportedness" in order to formalize this point, and examine the implications of the possibility of delegation with respect to well-supported equilibria.

Our second main result provides a complete answer to the query stated above by characterizing the conditions under which delegation obtains in *any* well-supported equilibrium.

Theorem 2. *There exists an* $\ell > 0$ *such that*

$$\beta_2[\emptyset](\neg D) = \begin{cases} 1, & \text{if } \Pi_2^{\text{SPE}} = \max \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) \\ 0, & \text{if } \Pi_2^{\text{SPE}} < \max \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) \end{cases}$$

for any well-supported equilibrium $(\beta, \mu) \in SE_{\text{w-s}}(\Lambda(\Gamma, \delta, c))$ *and any* $\delta + c < \ell$.

But does the presence of delegation imply that the pre-delegation outcome of the principals-only game will be altered? The answer is yes. An easy corollary of Theorems 1 and 2 is that, when delegation occurs, this always (strictly) benefits the delegating party, and hence the pre-delegation outcome (i.e., the subgame perfect equilibrium of Γ) is bound to be altered through delegation in a payoff-relevant way in *any* well-supported equilibrium.

Consequently, in a well-supported equilibrium, player 2 will choose not to delegate if he is already in an advantageous situation in the principals-only game Γ (in the sense that the *pre-delegation* outcome is already the best that he can achieve in any Nash equilibrium) whereas he will delegate if there is a Nash equilibrium in which he obtains a payoff strictly greater than his subgame perfect equilibrium payoff and in which player 1 plays sequentially rationally. So, for instance, in any (discrete) Stackelberg duopoly situation, the leader firm will not choose to delegate the decision-making power. On the other hand, by Theorem 2, the follower firm will (generically) choose to delegate the decision-making to an agent *even when the incentive contracts are fully unobservable.* Moreover, the delegation decision will certainly benefit the follower firm. Therefore, in sequential market games, it turns out that there is good reason to take Fersthman-Judd like delegation results seriously even in the presence of unobservable contracts.

Finally, to bring to the fore the basic intuitions behind these results, we now sketch the main steps of their proofs. The formal demonstrations are relegated to Section 7.

1. (Lemmas 1-2) The first step is to notice that any contract offered with positive probability in a sequential equilibrium of $\Lambda(\Gamma, \delta, c)$ is accepted with probability one and conditional on accepting such a contract agent A best responds to player 1's strategy according to the preferences of the principal. This is because (i) rejection leads to an expected payoff of

$\Pi_2^{\mathrm{SPE}} - c$, which is worse than the no-delegation payoff Π_2^{SPE}; (ii) if the contract does not lead the agent to best respond under that contract, then the principal can design an alternative (acceptable) contract that would force the agent best respond to player 1, thereby increasing the principal's expected payoff.

2. (Lemmas 3-5) The next step is to show that, given his beliefs about the contracts, the equilibrium behavior of player 1 in Λ after delegation, induces a sequentially rational strategy in the principals-only game Γ.

3. (Lemma 6) The above observations and the fact that player 2 would never delegate to obtain a payoff strictly smaller than his SPE payoff implies that $\mathbf{\Pi}_2^{\mathrm{SE}}(\Lambda) \subseteq \{\Pi_2 \in co\,\mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\mathrm{SPE}}\}$. Furthermore, for small enough cost of delegation, the converse containment holds as well, i.e., $\mathbf{\Pi}_2^{\mathrm{SE}}(\Lambda) \supseteq \{\Pi_2 \in \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\mathrm{SPE}}\}$.

4. (Lemma 7) If the cost of delegation is small enough, then player 2 never mixes between not delegating and offering a contract. This is because the set of gross payoffs that can be obtained by delegation is $\{\Pi_2 \in \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \Pi_2 > \Pi_2^{\mathrm{SPE}}\}$. So, if $\delta + c < \ell \equiv \min\{\alpha \in \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \alpha > \Pi_2^{\mathrm{SPE}}\} - \Pi_2^{\mathrm{SPE}}$, all equilibria involving delegation yields player 2 a net equilibrium payoff which is strictly greater than Π_2^{SPE}. Therefore, player 2 cannot be indifferent between delegating and not delegating in any sequential equilibria.

Steps (3) and (4) then imply that $\mathbf{\Pi}_2^{\mathrm{SE}}(\Lambda) = \{\Pi_2 \in \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\mathrm{SPE}}\}$ for small enough $\delta + c$. This establishes Theorem 1.

5. (Lemma 8) The next step is to prove that well-supportedness demands that player 2 chooses not to delegate only if there exists no equilibrium that involves delegation. To see this, let (β^*, μ^*) be a well-supported equilibrium in which player 2 chooses not to delegate with probability one (recall step (4)), and thus receives Π_2^{SPE}. Now suppose that there exists an equilibrium (β, μ) which involves delegation. From our earlier observations, (β, μ) must be yielding player 2 a net payoff strictly greater than Π_2^{SPE} for small $\delta + c$. Also, by definition of well-supportedness, the strategy of player 1 must be the same in β^* and β, following an information set, say I, that is reached by equilibrium β. One can show that player 2 can deviate in strategy profile β^* by offering a contract which, under β_A^*, makes the agent reach information set I with probability 1, and then play in a way such that the net payoff of player 2 is the same with his payoff under strategy profile β. In other words, there is a deviation for player 2 in equilibrium (β^*, μ^*) which gives him a payoff strictly greater than Π_2^{SPE}. This contradicts that (β^*, μ^*) is a sequential equilibrium.

Now observe that Theorem 1 implies that if $\Pi_2^{\mathrm{SPE}} = \max \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma)$, there cannot be any sequential equilibrium that involves delegation (since $\delta + c > 0$). On the other hand, for small enough $\delta + c$, if $\Pi_2^{\mathrm{SPE}} < \max \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma)$, then by Theorem 1, there exists a sequential equilibrium with delegation. Thus applying steps (4) and (5) yields $\beta_2[\emptyset](\neg D) = 0$, and establishes Theorem 2.

6 Caveats, Extensions and Concluding Remarks

Two-Sided Delegation. We assumed above that only one of the parties has the opportunity to delegate. A natural question, therefore, is if the findings reported here still have a "bite" if both of the principals have an opportunity to delegate. This problem is analyzed in Koçkesen (1999b) where it is assumed that prior to the game both principals can offer contracts to their delegates without being informed about whether the other party delegates or not. The party that plays the game (either the principal or the delegate) knows the identity (but not the contract) of her opponent in the game phase. The first result is that the set of sequential equilibrium outcomes of the two-sided delegation game is a subset of Nash equilibrium outcomes of the principals-only game, which is the counterpart of Proposition 1 in this paper.

However, making more precise predictions is possible only by considering a subset of sequential equilibria in which neither principal randomizes between delegating and offering a contract (they can, however, randomize between contracts). It is then possible to show that, in all such equilibria where only principal i delegates, the set of equilibrium gross payoffs is equal the set of Nash equilibrium payoffs of the principals-only game where principal i receives more than her SPE payoff, and the other party plays sequentially rationally. Furthermore, the set of equilibrium payoffs in which both parties delegate is equal to the set of Nash equilibrium payoffs in which both principals receive more than their individually rational payoffs (i.e., their minmax payoffs). This second result echoes our Theorem 1, and also indicates the potential use of delegation by the principals as a cooperative device to attain Pareto improvements over the subgame perfect equilibrium outcome.

As for the implications of well-supported equilibrium for delegation, Koçkesen (1999b) shows that only a weaker version of Theorem 2 applies to two-sided delegation games. Namely, if there exists no sequential equilibrium in which one or both principals randomize between delegating and not delegating, and if in the principals-only game any one of the principals can benefit by not playing in a sequentially rational manner, then *at least* one of the principals will choose to delegate in any well-supported equilibrium.

Renegotiation. An important assumption in our model is that contracts cannot be renegotiated once the outside party starts taking actions. This could be due to the physical impossibility of renegotiation, as it is the case in closed-door negotiations, or due to the fact that renegotiation is limited through a third party enforcement, or because it is too costly.[27]

In the framework of our paper, however, if costless renegotiation can take place at any point in the game and if the principal and the agent are symmetrically informed throughout the game, then delegation, with observed or unobserved contracts, would have no commitment power. This is because, due to the Pareto improving renegotiation opportunities, the delegate must behave sequentially rationally from the perspective of the delegating principal in any sequential equilibrium of such a delegation game. This may be contrasted with the work of Dewatripont (1988) and

[27] For instance, in the ultimatum bargaining game of Section 2, if renegotiation is as costly as the initial contract and $1 \leqslant c < 4 - \delta$, then delegation remains a sequential equilibrium even when renegotiation is allowed throughout the game.

Caillaud et al. (1995) who show that, if there is asymmetric information between the principal and the delegate at the time of contracting, publicly announced contracts may have a commitment value even if they are renegotiable *prior to the game stage.*[28]

Within the confines of the present framework, therefore, the positive results of this paper on the possibility of strategic delegation have applicability to real-world scenarios only to the extent that renegotiation is costly and/or limited.

Larger Classes of Principals-Only Games. While we have studied in Section 5 only those perfect information principals-only games with a unique subgame perfect equilibrium outcome, it is easy to generalize the present findings to larger classes of games. For instance, let Γ be any finite extensive form game with a unique sequential equilibrium outcome (and with $\left|\mathbf{\Pi}_2^{\mathrm{NE}}(\Gamma)\right| < \infty$). The proof of Theorem 2 modifies in a trivial manner to show that, whenever $\Pi_2^{\mathrm{SE}} < \max \mathbf{\Pi}_2^{\mathrm{NE}_1^*}(\Gamma)$ and $0 < \delta + c < \min\{\alpha \in \mathbf{\Pi}_2^{\mathrm{NE}}(\Gamma) : \alpha > \Pi_2^{\mathrm{SE}}\} - \Pi_2^{\mathrm{SE}}$, player 2 chooses to delegate with probability one in *all* well-supported equilibria of $\Lambda(\Gamma, \delta, c)$.

Another interesting extension of our model obtains by considering the simultaneous move principals-only games. As Corollary 1 and Theorem 2 (as extended in the previous paragraph) show, if Γ is a normal form game with a unique Nash equilibrium, then delegation never obtains in any equilibrium of the delegation game. However, as noted by a referee of this journal, if Γ has multiple Nash equilibria, although there may exist sequential equilibria with no delegation, forward induction type of refinements may lead to a selection in which the delegating principal obtains a strategic advantage. For example, if the principals-only game is that of the Battle of the Sexes game of Section 3 (but without the outside option), and if the cost of delegation is small enough, then all well-supported equilibria of the induced delegation game (which may or may not involve delegation) lead to the principal's preferred Nash outcome (T, L). This further emphasizes the two different dimensions of our results: first, if the underlying game is an extensive form game with a unique equilibrium outcome, it is possible that delegation would obtain as a sequential equilibrium of the induced delegation game; and, second, even if the underlying game is a simultaneous move game with multiple equilibria, the fact that delegation is a costly and observable decision may lead to equilibrium selection among the multiple Nash equilibria, provided that one subscribes to forward induction type refinements.

Principal-Agent Bargaining. Another aspect of the present model which could be fruitfully generalized is the bargaining process between the principal and the delegate. We assumed here that principal makes a "take it or leave it" offer to the delegate within a symmetric and complete information context. In reality, of course, these assumptions are rarely valid. An interesting conjecture is that the existence of asymmetric information between the principal and the delegate might restrict the contracts that would be offered in equilibrium in such a manner that Theorem 2 holds true under even weaker refinements of sequential equilibrium than what we have proposed. Furthermore,

[28] Since the principals-only games considered by these studies are simultaneous move games, the latter qualification is easily acceptable. In our model where the principals-only games have a sequential structure, however, this qualification is quite pressing.

the bargaining environment between the principal and the agent may impose limitations on the renegotiation possibilities later on, which may restore some commitment power to delegation even under renegotiation. The analysis of these issues, while certainly a promising avenue of research, falls outside the scope of the present paper.

Experiments. The main findings reported in this paper are based on a particular equilibrium refinement the empirical validity of which must be tested against the data. An obvious way to conduct this test is of course via experiments in which agents play a delegation game such as the simple bargaining game presented in Section 2. Fershtman and Gneezy (2001) conducts an experimental test of strategic delegation under observable and unobservable contracts within the context of such an ultimatum bargaining game. They provide some results which are not really in line with any of the theoretical results in the literature including those in Fershtman and Kalai (1997). We should note, however, that the structure of what Fershtman and Gneezy define as a delegation game is not the same as what we used in this paper. In particular, in their setting contract space is defined differently and the player who delegates does not have the option of playing the game himself. Consequently, testing the empirical validity of the theory we proposed here remains as an integral part of our future research agenda.

7 Proofs

A. Consistency and Sequential Rationality

Given an extensive form game $\Upsilon \equiv [N, H, P, (\mathcal{I}_i, \pi_i)_{i\in N}]$, a strategy profile β, and $h, h' \in H$, we let $p[\beta\,|\,h](h')$ be the probability of reaching history h', conditional on h being reached and from there on the game being played according to β. Also define $O[\beta\,|\,h]$ as the probability distribution over terminal nodes induced by β, conditional on h being reached. Given any strategy profile β, we define the *expected payoff of player i conditional on history h being reached* as

$$\Pi_i(\beta\,|\,h) \equiv \sum_{z\in Z} O[\beta\,|\,h](z)\pi_i(z).$$

To simplify the notation, we write $\Pi_i(\beta)$ for $\Pi_i(\beta\,|\,\emptyset)$, the expected payoff of player i induced by β in the entire game. The *expected payoff of player i conditional on his/her information set I* being reached is defined as

$$\Pi_i(\beta, \mu\,|\,I) \equiv \sum_{h\in I} \mu[I](h)\Pi_i(\beta\,|\,h).$$

In turn, an assessment (β, μ) is said to be *sequentially rational* if, for all $i \in N$ and all $I \in \mathcal{I}_i$,

$$\Pi_i(\beta, \mu\,|\,I) \geqslant \Pi_i((\beta_i', \beta_{-i}), \mu\,|\,I) \quad \text{for all } \beta_i' \in S_i(\Upsilon).$$

It is called *consistent* if there is a sequence of completely mixed assessments $((\beta^n, \mu^n))$, where each μ^n is derived from β^n by the Bayes rule, that converges to (β, μ). A consistent and sequentially rational assessment is called a *sequential equilibrium* (which is subject to the tie-breaking assumption mentioned in the final paragraph of Section 4 in the case of a delegation game).

B. Preliminary Observations

Consider a one-sided delegation environment $[\Gamma, \delta, c]$ and let $(\beta^*, \mu^*) \in SE(\Lambda(\Gamma, \delta, c))$. Clearly, for each $f \in \{0, \delta\}^Z$, the behavioral strategy β_A^* induces a behavioral strategy in the game Γ, which is defined as

$$b_{f,2}^*[h] \equiv \beta_A^*[f, a, h], \quad h \in H(2).$$

Similarly, β_1^* induces a behavioral strategy $b_1^* \in S_1(\Gamma)$, i.e.,

$$b_1^*[h] \equiv \beta_1^*[\{(f, a, h) : f \in \{0, \delta\}^Z\}], \quad h \in H(1).$$

For any $h \in H$, we denote by $o[b_1^*, b_{f,2}^* \,|\, h]$ the probability distribution over terminal nodes that will be reached if each player plays the game Γ according to the strategy profile $(b_1^*, b_{f,2}^*) \in S(\Gamma)$, conditional on h being reached. If $h = \emptyset$, we simply write $o[b_1^*, b_{f,2}^*]$ for $o[b_1^*, b_{f,2}^* \,|\, \emptyset]$. Consequently, we write the expected payoff of the agent who takes on a contract f by

$$F(b_1^*, b_{f,2}^*) \equiv \sum_{z \in Z} o[b_1^*, b_{f,2}^*](z) f(z).$$

(G is similarly defined for contract g.) The expected payoff of player 2 (gross of the payment to the agent and the contracting cost) is similarly written as $\Pi_2(b_1^*, b_{f,2}^*) \equiv \sum_{z \in Z} o[b_1^*, b_{f,2}^*](z)\pi_i(z)$. Of course, we have $\Pi_2^*(\beta^* \,|\, f) = p\left(\Pi_2(b_1^*, b_{f,2}^*) - F(b_1^*, b_{f,2}^*) - c\right) + (1-p)\left(\Pi_2^{\text{SPE}} - c\right)$ where $p = \beta_A^*[f](a)$, so that

$$\Pi_2^*(\beta^* \,|\, f) = \Pi_2^*(\beta^* \,|\, f, a) = \Pi_2(b_1^*, b_{f,2}^*) - F(b_1^*, b_{f,2}^*) - c \tag{2}$$

for any f with $\beta_A^*[f](a) = 1$.

C. Proof of Proposition 1

In what follows, we fix a one-sided delegation environment $[\Gamma, \delta, c]$, and denote the delegation game $\Lambda(\Gamma, \delta, c)$ by Λ for brevity. For any equilibrium strategy profile β^* in Λ, we let

$$C(\beta_2^*) \equiv supp(\beta_2^*[\emptyset]) \backslash \{\neg D\}.$$

Lemma 1. Let $(\beta^*, \mu^*) \in SE(\Lambda)$. For any $f \in C(\beta_2^*)$, we have

(*a*) $\Pi_2^*(\beta^* \,|\, f) \geqslant \Pi_2^{\text{SPE}}$,

(*b*) $\Pi_2^*(\beta^* \,|\, f) \geqslant \Pi_2^*(\beta^* \,|\, g)$ for all $g \in \{0, \delta\}^Z$ with equality for $g \in C(\beta_2^*)$,

(*c*) $\Pi_2^*(\beta^* \,|\, f) = \Pi_2^{\text{SPE}}$ if $\beta_2^*[\emptyset](\neg D) \in (0, 1)$.

Proof. Notice that $\Pi_2^*(\beta^* \,|\, \emptyset) = \beta_2^*[\emptyset](\neg D)\Pi_2^{\text{SPE}} + \sum_{g \in C(\beta_2^*)} \beta_2^*[\emptyset](g)\Pi_2^*(\beta^* \,|\, g)$ since $\Pi_2^*(\beta^* \,|\neg D) = \Pi_2^{\text{SPE}}$ by sequential rationality. But then, $\Pi_2^*(\beta^* \,|\, f) < \Pi_2^{\text{SPE}}$ and rationality at $\emptyset$ would entail that $\beta_2^*[\emptyset](f) = 0$ for any contract f, which proves part (a). Parts (b) and (c) are proved similarly. ☐

Lemma 2. Let $(\beta^*, \mu^*) \in SE(\Lambda)$. For any $f \in C(\beta_2^*)$, we have $\beta_A^*[f](a) = 1$, $F(b_1^*, b_{f,2}^*) = \delta$, and $b_{f,2}^* \in BR_2(b_1^*)$.[29]

[29] As usual, BR_i stands for the *best response* correspondence of player i in the principals-only game Γ.

Proof. Fix an arbitrary $f \in C(\beta_2^*)$ and take any strategy $b_2^* \in BR_2(b_1^*)$. Consider now the function $g \in \{0, \delta\}^Z$ defined as

$$g(z) \equiv \begin{cases} \delta, & \text{if } z \in supp(o[b_1^*, s_2]) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $s_2 \in supp(b_{f,2}^*)$.[30] We proceed by means of claims.

Claim 2.1. $b_{g,2}^* \in BR_2(b_1^*)$, $G(b_1^*, b_{g,2}^*) = \delta$, and $\beta_A^*[g](a) = 1$.

Proof of Claim 2.1. By definition of g, given μ^* and β_1^*, the agent must choose $b_{g,2}^*$ in order to guarantee that

$$supp(o[b_1^*, b_{g,2}^*]) \subseteq supp(o[b_1^*, s_2]). \tag{4}$$

This yields $G(b_1^*, b_{g,2}^*) = \delta$ and $\beta_A^*[g](a) = 1$ (since the agent is assumed to accept any contract that pays her δ).

Take any $s_2' \in supp(b_{g,2}^*)$. If $supp(o[b_1^*, s_2]) \subseteq supp(o[b_1^*, s_2'])$ did not hold, this would mean that there exists a history $h \in H(2)$ that is reached with positive probability under both (b_1^*, s_2) and (b_1^*, s_2') and that satisfies $s_2(h) \neq s_2'(h)$. But then it is obvious that this would imply $supp(o[b_1^*, s_2]) \cap supp(o[b_1^*, s_2']) = \emptyset$ which contradicts (4). Consequently, we have $supp(o[b_1^*, s_2]) \subseteq supp(o[b_1^*, s_2'])$. Since the converse containment follows from (4), we find $o[b_1^*, s_2](z) = o[b_1^*, s_2'](z)$ for all $z \in Z$. Therefore, since $s_2 \in BR_2(b_1^*)$, we must also have $s_2' \in BR_2(b_1^*)$. Given that s_2' is arbitrary in $supp(b_{g,2}^*)$, this establishes that $b_{g,2}^* \in BR_2(b_1^*)$. ‖

Claim 2.2. $\beta_A^*[f](a) = 1$.

Proof of Claim 2.2. By the tie-breaking assumption, $\beta_A^*[f](a) \in \{0, 1\}$. But if $\beta_A^*[f](a) = 0$ was the case, we would have $\Pi_2^*(\beta^* \mid f) = \Pi_2^{\text{SPE}} - c < \Pi_2^{\text{SPE}}$ which contradicts Lemma 1(a). ‖

Claim 2.3. $F(b_1^*, b_{f,2}^*) = \delta$.

Proof of Claim 2.3. Since Claim 2.2 shows that the agent accepts the contract f with probability one, her expected payoff conditional on accepting this contract must be equal to her outside option, that is, $F(b_1^*, b_{f,2}^*) \geqslant \delta$. The claim then follows from the fact that the highest possible expected payoff of the agent is δ. ‖

Our objective now is to establish that $\Pi_2(b_1^*, b_{f,2}^*) \geqslant \Pi_2(b_1^*, b_2)$ for all $b_2 \in S_2(\Gamma)$. To derive a contradiction, we assume that $\Pi_2(b_1^*, b_{f,2}^*) < \Pi_2(b_1^*, b_2)$ for some $b_2 \in S_2(\Gamma)$. Then, by Claim 2.1, we have $\Pi_2(b_1^*, b_{f,2}^*) < \Pi_2(b_1^*, b_{g,2}^*)$. But then using Claims 2.1-3, we find

$$\Pi_2^*(\beta^* \mid f) - \Pi_2^*(\beta^* \mid g) = (\Pi_2(b_1^*, b_{f,2}^*) - \Pi_2(b_1^*, b_{g,2}^*)) + (\delta - F(b_1^*, b_{f,2}^*)) < 0$$

which contradicts Lemma 1(b). The proof of Lemma 2 is then complete. □

Lemma 3. Let $(\beta^*, \mu^*) \in SE(\Lambda)$, and let $(f', a, z) \in supp(O[\beta^*])$. There exists a Nash equilibrium $\hat{b} \in NE(\Gamma)$ such that $z \in supp(o[\hat{b}])$ and $\hat{b}_1$ is a best response to some $b_2 \in S_2(\Gamma)$ at any history $h \in H(1)$.

[30] As is obvious, $s_2 \in supp(b_{f,2}^*)$ means that s_2 is a pure strategy of player 2 such that $s_2[h](a) = 1$ only if $b_{f,2}^*[h](a) > 0$ for all $h \in H(2)$.

Proof. If $C(\beta_2^*) = \emptyset$ then the claim is trivially established, so throughout we assume that $C(\beta_2^*) \neq \emptyset$. Let $H_o(1)$ stand for the set of all shortest histories in $H(1)$.[31] We let $I^h \equiv \{(f, a, h) : f \in \{0, \delta\}^Z\}$ for any $h \in H(1)$, and define

$$T^h \equiv \sum_{(f,a,h) \in I^h \cap supp(p[\beta^*|\emptyset])} p[\beta^*|\emptyset](f, a, h) \quad \text{for all } h \in H_o(1),$$

It is important to note that the definition of T^h is independent of β_1^* since $p[\beta^*|\emptyset](f, a, h)$ is independent of β_1^* for any shortest history h in $H(1)$. Therefore, by weak consistency and sequential rationality, we have

$$\beta_1^* \in \arg\max_{\beta_1 \in S_1(\Lambda)} \sum_{\substack{f \in C(\beta_2^*) \\ z \in Z'}} \Big(\mu^*[I^h](f, a, h)\Big)\, O[\beta_1, \beta_{-1}^* \,|\, f, a, h](z)\pi_1(z)$$

for all $h \in H_o(1)$ with $p[\beta^*|\emptyset](f, a, h) > 0$ for some f, where $Z' = supp(O[\beta_1, \beta_{-1}^* \,|\, f, a, h])$. Recalling that $b_1^* \in S_1(\Gamma)$ is the behavioral strategy in Γ induced by β_1^*, we thus have, for all $h \in H_o(1)$ with $T^h > 0$,

$$\begin{aligned} b_1^* &\in \arg\max_{b_1 \in S_1(\Gamma)} \sum \Big(\mu^*[I^h](f, a, h)\Big)\, o[b_1, b_{f,2}^* \,|\, h](z)\pi_1(z) \\ &= \arg\max_{b_1 \in S_1(\Gamma)} \sum \tfrac{1}{T^h} p[\beta^*|\emptyset](f, a, h) o[b_1, b_{f,2}^* \,|\, h](z)\pi_1(z) \\ &= \arg\max_{b_1 \in S_1(\Gamma)} \sum \beta_2^*[\emptyset](f) p[\beta^*|f](f, a, h) o[b_1, b_{f,2}^* \,|\, h](z)\pi_1(z) \end{aligned} \tag{5}$$

where the sums run through all $f \in C(\beta_2^*)$ and $z \in Z$. (Recall that, for all such h, consistency assures that $\mu^*[I^h](f, a, h) = p[\beta^*|\emptyset](f, a, h)/T^h$ and $\beta_A^*[f](a) = 1$ for all $f \in C(\beta_2^*)$ by Lemma 2.) But notice that (5) holds trivially for any $h \in H_o(1)$ such that $p[\beta^*|\emptyset](f, a, h) = 0$ for all $f \in \{0, \delta\}^Z$, since in this case the maximand of the associated optimization problem is identically zero. Consequently, (5) holds for all $h \in H_o(1)$, and we thus have

$$b_1^* \in \arg\max_{b_1 \in S_1(\Gamma)} \sum_{\substack{f \in C(\beta_2^*) \\ z \in Z}} \beta_2^*[\emptyset](f) \left(\sum_{h \in H_o(1)} p[\beta^*|f](f, a, h) o[b_1, b_{f,2}^* \,|\, h](z) \right) \pi_1(z). \tag{6}$$

Claim 3.1. For all $z \in Z \cap \bigcup_{h \in H_o(1)} \mathcal{H}(h)$,

$$\sum_{h \in H_o(1)} p[\beta^*|f](f, a, h) o[b_1, b_{f,2}^* \,|\, h](z) = o[b_1, b_{f,2}^*](z).$$

Proof of Claim 3.1. Take any $z \in Z \cap \bigcup_{h \in H_o(1)} \mathcal{H}(h)$. By definition of $H_o(1)$, there can be at most one h in $H_o(1)$ that is consistent with z. By the choice of z, therefore, there exists a unique $h_z \in H_o(1)$ that is consistent with z. But then, $o[b_1, b_{f,2}^* \,|\, h](z) = 0$ for all $h \neq h_z$, so, by definition of $p[\beta^*|f]$,

$$\sum_{h \in H_o(1)} p[\beta^*|f](f, a, h) o[b_1, b_{f,2}^* \,|\, h](z) = p[\beta^*|f](f, a, h_z) o[b_1, b_{f,2}^* \,|\, h_z](z) = o[b_1, b_{f,2}^*](z). \;\|$$

[31] That is, $H_o(1) \equiv \{h \in H(1) : \text{there does not exist any } h' \in H(1) \text{ and } h'' \neq \emptyset \text{ such that } h = (h', h'')\}$.

Using Claim 3.1 and (6), we find

$$b_1^* \in \arg\max_{b_1 \in S_1(\Gamma)} \sum_{\substack{f \in C(\beta_2^*) \\ z \in Z}} \beta_2^*[\emptyset](f) o[b_1, b_{f,2}^*](z)\pi_1(z)$$

since it is readily observed that, for any f, the probability $o[b_1, b_{f,2}^*](z)$ is independent of b_1 for any terminal history $z \notin \bigcup_{h \in H_o(1)} \mathcal{H}(h)$. Now let $\lambda(f) = \dfrac{\beta_2[\emptyset](f)}{1 - \beta_2[\emptyset](\neg D)}$ for all $f \in C(\beta_2^*)$. We have $\sum_{f \in C(\beta_2^*)} \lambda(f) = 1$ and

$$b_1^* \in \arg\max_{b_1 \in S_1(\Gamma)} \sum_{\substack{f \in C(\beta_2^*) \\ z \in Z}} \lambda(f) o[b_1, b_{f,2}^*](z)\pi_1(z). \tag{7}$$

Moreover, for any $b_1 \in S_1(\Gamma)$,

$$\sum_{\substack{f \in C(\beta_2^*) \\ z \in Z}} \lambda(f) o[b_1, b_{f,2}^*](z)\pi_1(z) = \sum_{z \in Z} o\left[b_1, \sum_{f \in C(\beta_2^*)} \lambda(f) b_{f,2}^*\right](z)\pi_1(z) = \Pi_1\left(b_1, \sum_{f \in C(\beta_2^*)} \lambda(f) b_{f,2}^*\right)$$

so that, by (7), we have $b_1^* \in \arg\max_{b_1 \in S_1(\Gamma)} \Pi_1(b_1, \hat{b}_2)$, and hence $b_1^* \in BR_1(\hat{b}_2)$ where $\hat{b}_2 \in S_2(\Gamma)$ is defined as $\hat{b}_2 \equiv \sum_{f \in C(\beta_2^*)} \lambda(f) b_{f,2}^*$. But since, by Lemma 2, $b_{f,2}^* \in BR_2(b_1^*)$ for all $f \in C(\beta_2^*)$, we also have $\hat{b}_2 \in BR_2(b_1^*)$. Therefore, $(b_1^*, \hat{b}_2) \in NE(\Gamma)$. Furthermore, sequential rationality of player 1 at any I^h, $h \in H(1)$, implies that b_1^* is a best response to some $b_2 \in S_2(\Gamma)$ with $b_2 = \sum b_{f,2}^* \mu[I^h](f)$ for some probability measure μ on $\{0, \delta\}^Z$. Moreover, if $(f', a, z) \in supp(O[\beta^*])$, then it must be the case that $f' \in C(\beta_2^*)$ and $z \in supp(o[b_1^*, b_{f',2}^*])$, and hence $z \in supp(o[b_1^*, \hat{b}_2])$. □

To state the next lemma, let $h \in H \backslash Z$ be any non-terminal history, and define $\mathcal{H}(h)$ as the set of all histories consistent with h.[32]

Lemma 4. Let $b_1 \in S_1(\Gamma)$ and let $B_2^h(b_1)$ be the set of all $b_2 \in S_2(\Gamma)$ such that $b_1 \in BR_1(b_2 \mid h)$ at an history $h \in H(1)$. If $B_2^h(b_1) \neq \emptyset$ for all $h \in H(1)$, then

$$\bigcap \{B_2^h(b_1) : h \in \mathcal{H}(h^*) \cap H(1)\} \neq \emptyset \quad \text{for any } h^* \in H(1).$$

Proof. Fix a $b_1 \in S_1(\Gamma)$ and a $h^* \in H(1)$. Since $\mathcal{H}(h^*) \cap H(1)$ is obviously a non-empty finite set, we may write, for some positive integer n, $\mathcal{H}(h^*) \cap H(1) = \{h_1, h_2, \ldots, h_n\}$ where $|h_1| \leqslant \cdots \leqslant |h_n|$. We will prove by induction that $\bigcap \{B_2^{h_j}(b_1) : n \geqslant j \geqslant k\} \neq \emptyset$ for all $k = 1, 2, \ldots, n$.

The statement is trivially true for $k = n$. Now, suppose that it is true for $k = l \leqslant n$. Let $b_2^* \in B_2^{h_{l-1}}(b_1)$ and $\bar{b}_2 \in B_2^{h_n}(b_1) \cap \cdots \cap B_2^{h_l}(b_1)$. Also let $H^k = \{h_k, h_{k+1}, \ldots, h_n\}$, for all $k \in \{1, 2, \ldots, n\}$. If (b_1, b_2^*) reaches every history in H^l with positive probability, then b_1 must be a best response to b_2^* at every history in H^l as well, so $b_2^* \in \bigcap \{B_2^{h_j}(b_1) : n \geqslant j \geqslant l-1\}$, and we are done. Suppose, therefore, that there exists a history in H^l which is precluded by (b_1, b_2^*). Let h_t be the shortest such history. Define

$$b_2'[h] = \begin{cases} \bar{b}_2[h], & \text{if } h \in \mathcal{H}(h_t) \cap H(2) \\ b_2^*[h], & \text{if } h \in H(2) \backslash \mathcal{H}(h_t) \end{cases}.$$

[32] Formally speaking, $h'' \in \mathcal{H}(h)$ iff $h'' = (h, h')$ for some h'.

Since $b_1 \in BR_1(b_2^* \,|\, h_{l-1})$, the probability of reaching h_t under (b_1, b_2^*) is zero, and all $h \in \{h_{l-1}, \ldots, h_{t-1}\}$ are reached with positive probability, we have $b_1 \in BR_1(b_2' \,|\, h)$ for all such h, which means that $b_2' \in B_2^{h_{t-1}}(b_1) \cap \cdots \cap B_2^{h_{l-1}}(b_1)$. But, by induction hypothesis, we have $b_2' \in B_2^{h_n}(b_1) \cap \cdots \cap B_2^{h_t}(b_1)$, so combining these observations we again get $b_2^* \in \bigcap\{B_2^{h_j}(b_1) : n \geqslant j \geqslant l-1\}$. □

Lemma 5. Let $(b_1^*, b_2^*) \in NE(\Gamma)$. If, for any $h \in H(1)$,

$$b_1^* \in \arg\max_{b_1 \in S_1(\Gamma)} \Pi_1(b_1, b_2 \,|\, h) \quad \text{ for some } b_2 \in S_2(\Gamma),$$

then there exists a behavioral strategy profile $(b_1', b_2') \in NE_1^*(\Gamma)$ such that $o[b_1', b_2'] = o[b_1^*, b_2^*]$.

Proof. If $(b_1^*, b_2^*) \in NE_1^*(\Gamma)$ there is nothing to prove. So, suppose that b_1^* is not a best response to b_2^* at some history. Since $(b_1^*, b_2^*) \in NE(\Gamma)$, any such history must be off-the-path of (b_1^*, b_2^*). Let $\hat{H} = \{h_1, h_2, \ldots, h_n\}$ denote the set of all such histories. Clearly, for each $h_j \in \hat{H}$, there exists an earlier history $h \in H$ such that, when $b_{P(h)}^*[h]$ is played, h_j can never be reached, no matter how the behavioral strategies at other histories are specified. Let $\bar{h}_j$ denote the shortest one among all those histories. Notice that since there is a unique sequence of actions that reach h_j, the history $\bar{h}_j$ is uniquely defined.

Partition $\hat{H}$ into $m \leqslant n$ disjoint sets, $\hat{H}_\alpha$, $\alpha = 1, \ldots m$, such that two histories h_j and h_k belong to the same set $\hat{H}_\alpha$ if and only if $\bar{h}_j = \bar{h}_k$. Let $\hat{H}_\alpha^{\min}$ be the set of the shortest histories in each $\hat{H}_\alpha$. We will now construct a strategy profile $(b_1', b_2') \in NE_1^*(\Gamma)$ with $o[b_1', b_2'] = o[b_1^*, b_2^*]$. To do this, fix an α, and pick any $h \in \hat{H}_\alpha$. Assume first that $P(\bar{h}) = 1$. Then, change b_1^* at each history in $\hat{H}_\alpha \cap H(1)$ so as to make the new strategy a best response to b_2^* at those histories. (We can do this by applying a simple backward induction in every subgame starting with a history in $\hat{H}_\alpha^{\min}$). Assume next that $P(\bar{h}) = 2$. In this case, since b_1^* is a best response to some b_2 at any history, we can use Lemma 4 to change b_2^* to make b_1^* a best response to it at all histories in $\hat{H}_\alpha \cap H(1)$. Leave (b_1^*, b_2^*) unchanged in all histories which do not belong to $\hat{H}$ and call the newly constructed strategy profile (b_1', b_2').

By construction, b_1' is a best response to b_2' at every history. Furthermore, since (b_1', b_2') is different from (b_1^*, b_2^*) at only those histories which are not reached by (b_1^*, b_2^*) we have $o[b_1', b_2'] = o[b_1^*, b_2^*]$. Therefore, if we can show that $(b_1', b_2') \in NE(\Gamma)$, we will be done. To do this, notice that, since b_2' is different from b_2^* at only those histories excluded by b_2^* we have $\Pi_i(b_1, b_2^*) = \Pi_i(b_1, b_2')$ for all $b_1 \in S_1(\Gamma)$, $i = 1, 2$. Therefore, given that $(b_1^*, b_2^*) \in NE(\Gamma)$, we have $\Pi_1(b_1', b_2') = \Pi_1(b_1^*, b_2^*) \geqslant \Pi_1(b_1, b_2^*) = \Pi_1(b_1, b_2')$ for all $b_1 \in S_1(\Gamma)$. That $\Pi_2(b_1', b_2') \geqslant \Pi_1(b_1', b_2)$ holds for all $b_2 \in S_2(\Gamma)$ is shown in the analogous way, and this completes the proof. □

The proof of Proposition 1 is now completed upon applying Lemmas 3 and 5.

D. Proofs of Theorems 1, 2 and Corollary 2

Define

$$\mathbf{\Pi}_2^{\text{SE}}(\Lambda) \equiv \{\tilde{\Pi}_2(\beta) : (\beta, \mu) \in SE(\Lambda) \text{ for some system of beliefs } \mu\},$$

where

$$\tilde{\Pi}_2(\beta) \equiv \beta_2[\emptyset](\neg D)\Pi_2^{\text{SPE}} + \sum_{f\in C(\beta_2)} \beta_2[\emptyset](f)\Pi_2(b_1, b_{f,2}) \tag{8}$$

for any behavioral strategy profile $\beta \in S(\Lambda)$. Since, by Lemma 2, all contracts $f \in C(\beta_2^*)$ are accepted in equilibrium, $\tilde{\Pi}_2(\beta)$ is the expected payoff of player 2 (gross of the compensation he pays to the delegate in case of a hire and the contracting cost) in the equilibrium (β, μ). The following important lemma points to the close connection between the sets $\mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$ and $\mathbf{\Pi}_2^{\text{SE}}(\Lambda)$, and constitutes a crucial step towards proving Theorem 1.

Lemma 6. ***(a)*** $\mathbf{\Pi}_2^{\text{SE}}(\Lambda) \subseteq \{\Pi_2 \in co\,\mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\text{SPE}}\}$.

(b) There exists an $\ell > 0$ such that, for all $\delta + c < \ell$,

$$\mathbf{\Pi}_2^{\text{SE}}(\Lambda) \supseteq \{\Pi_2 \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\text{SPE}}\}.$$

Proof. ***(a)*** Let $\Pi_2 \in \mathbf{\Pi}_2^{\text{SE}}(\Lambda)$ and take any $(\beta^*, \mu^*) \in SE(\Lambda)$ such that $\tilde{\Pi}_2(\beta^*) = \Pi_2$. By Lemma 1(a), we have $\Pi_2(b_1^*, b_{f,2}^*) > \Pi_2^*(\beta^* \mid f) \geqslant \Pi_2^{\text{SPE}}$ so that $\tilde{\Pi}_2(\beta^*) \geqslant \Pi_2^{\text{SPE}}$, and it remains to show that $\Pi_2 \in co\ \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$. Notice first that if $\beta_2^*[\emptyset](\neg D) = 1$, then $\Pi_2 = \tilde{\Pi}_2(\beta^*) = \Pi_2^{\text{SPE}} \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, so we are done. If, on the other hand, $\beta_2^*[\emptyset](\neg D) = 0$, we define $\hat{b}$ as in the proof of Lemma 3, so that

$$\hat{b}_2 \equiv \sum_{f\in C(\beta_2^*)} \beta_2^*[\emptyset](f) b_{f,2}^*.$$

Recall that $(b_1^*, \hat{b}_2) \in NE(\Gamma)$, and, by Lemma 5, there exists a $(b_1', b_2') \in NE_1^*(\Gamma)$ with $o[b_1', b_2'] = o[b_1^*, \hat{b}_2]$. Therefore, we have

$$\tilde{\Pi}_2(\beta^*) = \sum_{f\in C(\beta_2^*)} \beta_2^*[\emptyset](f)\Pi_2(b_1^*, b_{f,2}^*) = \Pi_2\left(b_1^*, \sum_{f\in C(\beta_2^*)} \beta_2^*[\emptyset](f) b_{f,2}^*\right) = \Pi_2(b_1^*, \hat{b}_2) = \Pi_2(b_1', b_2') \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$$

and the claim follows.

To complete the proof we need to consider the case in which $\beta_2^*[\emptyset](\neg D) \in (0, 1)$. But this case is easily settled by noticing that

$$\begin{aligned}
\tilde{\Pi}_2(\beta^*) &= \beta_2^*[\emptyset](\neg D)\Pi_2^{\text{SPE}} + (1 - \beta_2^*[\emptyset](\neg D))\Pi_2\left(b_1^*, \textstyle\sum_{f\in C(\beta_2^*)} \beta_2^*[\emptyset](f) b_{f,2}^*\right) \\
&= \beta_2^*[\emptyset](\neg D)\Pi_2^{\text{SPE}} + (1 - \beta_2^*[\emptyset](\neg D))\Pi_2(b_1^*, \hat{b}_2) \\
&= \beta_2^*[\emptyset](\neg D)\Pi_2^{\text{SPE}} + (1 - \beta_2^*[\emptyset](\neg D))\Pi_2(b_1', b_2') \\
&\in co\mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma),
\end{aligned}$$

where we again used the bilinearity of Π_2 and Lemma 5.

(b) Let $\Pi_2 \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$ and $\Pi_2 \geqslant \Pi_2^{\text{SPE}}$. We need to show that there exists a $(\beta, \mu) \in SE(\Lambda)$ such that $\tilde{\Pi}_2(\beta) = \Pi_2$. We first introduce some notation and then construct such an equilibrium by distinguishing between two cases.

[*Case 1.* $\Pi_2 = \Pi_2^{\text{SPE}}$] Let $b^{\text{SPE}} \in SPE(\Gamma)$, and, for each $f \in \{0,\delta\}^Z$, take any $b^f \in S_2(\Gamma)$ such that $b^f[h] = b^{f,h}[h]$ for all $h \in H(2)$, where

$$b^{f,h} \in \arg\max_{b_2 \in S_2(\Gamma)} \sum_{z\in Z} o[b_1^{\text{SPE}}, b_2 \,|\, h](z) f(z).$$

Let $g \equiv \delta \mathbf{1}_{supp(o[b^{\text{SPE}}]))}$, and define the behavioral strategy profile $\beta \in S(\Lambda)$ as follows: For all $f \in \{0,\delta\}^Z$,

$$\beta_1[\neg D, h] = \beta_1[f,a,h] = \beta_1[f,r,h] = b_1^{\text{SPE}}[h] \text{ for all } h \in H(1),$$

$$\beta_2[\emptyset](\neg D) = 1,\ \beta_2[\neg D, h] = \beta_2[f,r,h] = b_2^{\text{SPE}}[h] \text{ for all } h \in H(2),$$

$$\beta_A[f](a) = \begin{cases} 1, & \text{if } F(b_1^{\text{SPE}}, b^f) = \delta \\ 0, & \text{otherwise} \end{cases},\ \beta_A[g](a) = 1,$$

$$\beta_A[f,a,h] = \begin{cases} b_2^{\text{SPE}}[h], & \text{if } f = g \\ b^f[h], & \text{otherwise} \end{cases} \text{ for all } h \in H(2),$$

On the other hand, we define $\mu \in B(\Lambda)$ by specifying that $\mu[I](g,\theta,h) = 1$ for all $I = \{(f,\theta,h) : f \in \{0,\delta\}^Z\} \in \mathcal{I}_1^*$, $\theta \in \{a,r\}$.

Sequential rationality of (β,μ) easily follows. To prove consistency, consider the completely mixed assessments $(\beta^\varepsilon,\mu^\varepsilon)$ defined as follows: Let $h_{\max}$ be the longest history in Γ, and

$$\beta_2^\varepsilon[\emptyset](g) = \varepsilon,\ \beta_2^\varepsilon[\emptyset](f) = \varepsilon^{|h_{\max}|+2},\ \text{for all } f \in \{0,\delta\}^Z \setminus \{g\}.$$

For all $i \in \{1,2,A\}$, $f \in \{0,\delta\}^Z$, $\theta \in \{a,r\}$, and $I = \{(f,\theta,h) : f \in \{0,\delta\}^Z\} \in \mathcal{I}_1^*$,

$$\beta_i^\varepsilon[f,\theta,h](x) = \begin{cases} \beta_i[f,\theta,h](x)\left[1 - \left(|A(h)| - |supp(\beta_i[f,\theta,h])|\right)\varepsilon\right], & \text{if } x \in supp(\beta_i[f,\theta,h]) \\ \varepsilon, & \text{otherwise} \end{cases},$$

and

$$\mu^\varepsilon[I](f,\theta,h) = \frac{p[\beta^\varepsilon \,|\, \emptyset](f,\theta,h)}{\sum_{f' \in \{0,\delta\}^Z} p[\beta^\varepsilon \,|\, \emptyset](f',\theta,h)}.$$

Clearly, $\beta^\varepsilon \to \beta$ as $\varepsilon \to 0$, and μ^ε is derived from β^ε using Bayes' rule. Moreover,

$$\begin{aligned} \mu^\varepsilon[I](g,\theta,h) &= \frac{p[\beta^\varepsilon \,|\, \emptyset](g,\theta,h)}{p[\beta^\varepsilon \,|\, \emptyset](g,\theta,h) + \sum_{f \in \{0,\delta\}^Z \setminus \{g\}} p[\beta^\varepsilon \,|\, \emptyset](f,\theta,h)} \\ &\geqslant \frac{\varepsilon(1 - (|\{0,\delta\}^Z| - 1)\varepsilon)\varepsilon^{|h_{\max}|}}{\varepsilon(1 - (|\{0,\delta\}^Z| - 1)\varepsilon)\varepsilon^{|h_{\max}|} + \varepsilon^{|h_{\max}|+2}(1-\varepsilon)(|\{0,\delta\}^Z| - 1)} \\ &\to 1 \end{aligned}$$

as $\varepsilon \to 0$.[33]

[33] What makes (β,μ) an equilibrium is the fact that in this assessment player 1 believes that player 2 has aligned the incentives of A with his own perfectly at all out-of-equilibrium information sets.

[*Case 2.* $\Pi_2 > \Pi_2^{\text{SPE}}$] Since $\Pi_2 \in \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, there exists a $\hat{b} \in NE(\Gamma)$ such that $\Pi_2(\hat{b}) = \Pi_2$, and $\hat{b}_1$ is sequentially rational. Let $g \equiv \delta \mathbf{1}_{supp(o[\hat{b}])}$, and define the behavioral strategy profile $\beta \in S(\Lambda)$ as follows: For all $f \in \{0, \delta\}^Z$,

$$\beta_1[\neg D, h] = \beta_1[f, r, h] = b_1^{\text{SPE}}[h],\ \beta_1[f, a, h] = \hat{b}_1[h] \quad \text{for all } h \in H(1),$$

$$\beta_2[\emptyset](g) = 1,\ \beta_2[\neg D, h] = \beta_2[f, r, h] = b_2^{\text{SPE}}[h] \ \text{ for all } h \in H(2),$$

$$\beta_A[f](a) = \begin{cases} 1, & \text{if } F(\hat{b}_1, b^f) = \delta \\ 0, & \text{otherwise} \end{cases},\ \beta_A[g](a) = 1,$$

$$\beta_A[f, a, h] = \begin{cases} \hat{b}_2[h], & \text{if } f = g \\ b^f[h], & \text{otherwise} \end{cases} \quad \text{for all } h \in H(2),$$

with $b^f \in S_2(\Gamma)$ being such that $b^f[h] = b^{f,h}[h]$ for all $h \in H(2)$, where

$$b^{f,h} \in \arg\max_{b_2 \in S_2(\Gamma)} \sum_{z \in Z} o[\hat{b}_1, b_2 \,|\, h](z) f(z).$$

We define $\mu \in B(\Lambda)$ by specifying that $\mu[I](g, \theta, h) = 1$ for all $I = \{(f, \theta, h) : f \in \{0, \delta\}^Z\} \in \mathcal{I}_1^*$, $\theta \in \{a, r\}$. Proof of sequential rationality and consistency of (β, μ) is similar to Case 1 above. $\square$

Lemma 7. If $\Pi_2^{\text{SPE}} = \max \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, then $\beta_2[\emptyset](\neg D) = 1$. Moreover, there exists an $\ell > 0$ such that $\beta_2[\emptyset](\neg D) \in \{0, 1\}$ for any $(\beta, \mu) \in SE(\Lambda)$ and any $\delta + c < \ell$.

Proof. Fix any $(\beta^*, \mu^*) \in SE(\Lambda)$ and notice that Lemma 1(b) and (2) imply that $\Pi_2(b_1^*, b_{f,2}^*) = \Pi_2(b_1^*, b_{g,2}^*)$ for all $f, g \in C(\beta_2^*)$. Therefore, by (8),

$$\tilde{\Pi}_2(\beta^*) = \beta_2^*[\emptyset](\neg D)\Pi_2^{\text{SPE}} + (1 - \beta_2^*[\emptyset](\neg D))\Pi_2(b_1^*, b_{f,2}^*) \tag{9}$$

for any $f \in C(\beta_2^*)$. Now if $\Pi_2^{\text{SPE}} = \max \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, then by Lemma 6(a), $\tilde{\Pi}_2(\beta^*) = \Pi_2^{\text{SPE}}$ and (9) yields $\Pi_2(b_1^*, b_{f,2}^*) = \Pi_2^{\text{SPE}}$ for any $f \in C(\beta_2^*)$. But then, by (2) and Lemma 2, $\Pi_2^*(\beta^* \,|\, f) = \Pi_2(b_1^*, b_{f,2}^*) - \delta - c < \Pi_2^{\text{SPE}}$ for any $f \in C(\beta_2^*)$. In view of Lemma 1(a), this can hold only if $C(\beta_2^*) = \{\neg D\}$, which implies that $\beta_2^*(\emptyset)(\neg D) = 1$.

Assume next that $\Pi_2 > \Pi_2^{\text{SPE}}$ for some $\Pi_2 \in \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, and define

$$\ell \equiv \min \left\{ \Pi_2 - \Pi_2^{\text{SPE}} : \Pi_2^{\text{SPE}} < \Pi_2 \in \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma) \right\}.$$

(That ℓ is well-defined follows from the hypothesis that $\left|\boldsymbol{\Pi}_2^{\text{NE}}\right| < \infty$.) To derive a contradiction, assume next that $\ell > \delta + c$ and $\beta_2^*[\emptyset](\neg D) \in (0, 1)$. Then, by Lemma 1(c), $\Pi_2^*(\beta^* \,|\, f) = \Pi_2^{\text{SPE}}$ so that, by (2) and Lemma 2, $\Pi_2(b_1^*, b_{f,2}^*) = \Pi_2^{\text{SPE}} + \delta + c$ for all $f \in C(\beta_2^*)$. Defining λ and $\hat{b}_2$ as in the proof of Lemma 3, we thus find

$$\Pi_2(b_1^*, \hat{b}_2) = \Pi_2 \left(b_1^*, \sum_{f \in C(\beta_2^*)} \lambda(f) b_{f,2}^* \right) = \sum_{f \in C(\beta_2^*)} \lambda(f) \Pi_2(b_1^*, b_{f,2}^*) = \Pi_2^{\text{SPE}} + \delta + c.$$

Lemma 5 implies that there exists a behavioral strategy profile $(b_1', b_2') \in NE_1^*(\Gamma)$ with $o[b_1', b_2'] = o[b_1^*, \hat{b}_2]$, and with $\Pi_2(b_1', b_2') = \Pi_2^{\text{SPE}} + \delta + c$. But then $\Pi_2^{\text{SPE}} + \delta + c \in \{\Pi_2 \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) : \Pi_2 > \Pi_2^{\text{SPE}}\}$, and thus by definition of ℓ, we obtain $\ell \leqslant \Pi_2^{\text{SPE}} + \delta + c - \Pi_2^{\text{SPE}} = \delta + c$. □

Proof of Theorem 1. From Lemma 7 and the first paragraph of the proof of Lemma 6(a), it follows that there exists an $\ell > 0$ such that $\mathbf{\Pi}_2^{\text{SE}}(\Lambda) \subseteq \{\Pi_2 \in \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma) : \Pi_2 \geqslant \Pi_2^{\text{SPE}}\}$ for all any $\delta + c < \ell$. The claim then follows from Lemma 6(b). □

Lemma 8. Assume that $\Pi_2^{\text{SPE}} < \max \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$. There exists an $\ell > 0$ such that, whenever $\delta + c < \ell$, for any well-supported equilibrium $(\beta^*, \mu^*) \in SE(\Lambda(\Gamma, \delta, c))$, we have $\beta_2^*[\emptyset](\neg D) = 1$ if and only if there exists no $(\beta, \mu) \in SE(\Lambda(\Gamma, \delta, c))$ with $\beta_2[\emptyset](\neg D) < 1$.

Proof. Pick an arbitrary $(\beta^*, \mu^*) \in SE_{\text{w-s}}(\Lambda)$, choose ℓ as in Lemma 7, assume that $\delta + c < \ell$, and recall that $\beta_2^*[\emptyset](\neg D) \in \{0, 1\}$ by Lemma 7. So the "if" part of the claim is trivial. To prove the converse, let $\beta_2^*[\emptyset](\neg D) = 1$ and assume that there exists an equilibrium of Λ in which delegation obtains with probability 1. Since $\Pi_2^{\text{SPE}} < \max \mathbf{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, player 1 must move at least once on the equilibrium path, and this guarantees that player 1 has an information set $I \in \mathcal{J}(\beta^*)$ (recall Definition 1) visited with positive probability in an equilibrium. Thus, by well-supportedness of (β^*, μ^*), there exists an assessment $(\beta, \mu) \in SE(\Lambda; I)$ such that $\mu^*[J] = \mu[J]$ and

$$\beta_1^*[J] = \beta_1[J] \qquad \text{for all } J \in \mathcal{I}_1(I), \tag{10}$$

where $\mathcal{I}_1(I)$ is the set of all information sets of player 1 that follows I.

Let us now pick any $f \in C(\beta_2)$ such that $b_{f,2}$ reaches I with positive probability. We have

Claim 8.1. $\Pi_2^*(\beta \,|\, f) > \Pi_2^{\text{SPE}}$.

Proof of Claim 8.1. If $\Pi_2^*(\beta \,|\, f) = \Pi_2^{\text{SPE}}$ held, then we would have $((\beta_1, \beta_2', \beta_A), \mu) \in SE(\Lambda)$ where $\beta_2' \in S_2(\Lambda)$ is defined as $\beta_2'[\emptyset](\neg D) \equiv 1/2$, $\beta_2'[\emptyset](g) \equiv 2^{-1}\beta_2[\emptyset](g)$ for all $g \in \{0, \delta\}^Z$ and $\beta_2'[h] \equiv \beta_2[h]$ for all $h \in H^*(2)\backslash\{\emptyset\}$. This, however, contradicts Lemma 7. We must then have $\Pi_2^*(\beta \,|\, f) \neq \Pi_2^{\text{SPE}}$, and the claim follows from Lemma 1(a). ‖

By the choice of f, $b_{f,2}$ reaches I with positive probability. (Recall that player 1 does not play a role in I being reached since $I = I^h$ with h being a shortest history in $H(1)$). We choose any pure strategy $s_2 \in supp(b_{f,2})$ such that s_2 reaches I.[34] Define next the contract $g \equiv \delta \mathbf{1}_{supp(o[b_1, s_2])}$.

Claim 8.2. $G(b_1, s_2) = \delta = G(b_1^*, b_{g,2}^*)$, and hence $\beta_A^*[g](a) = 1$.

Proof of Claim 8.2. By definition of g and sequential rationality, any $s_2' \in supp(b_{g,2}^*)$ must reach I regardless of the strategy choice of player 1. Therefore, all information sets that are reached when $b_{g,2}^*$ is played follow I regardless of the play of 1. Then, by (10), we must have $o[b_1, b_{g,2}^*] = o[b_1^*, b_{g,2}^*]$. Since s_2 reaches I as well, we also have $o[b_1, s_2] = o[b_1^*, s_2]$. Thus by sequential rationality, we find $\delta \geqslant G(b_1^*, b_{g,2}^*) \geqslant G(b_1^*, s_2) = G(b_1, s_2) = \delta$. ‖

[34] Formally speaking, by "s_2 reaches I" we mean that I is visited with positive probability via any strategy profile in Λ such that f is played with positive probability and the agent plays according to s_2 upon acceptance of f (recall Lemma 2). Also notice that since I is the "first" information set, s_2 reaches I with probability 1.

Claim 8.3. $\Pi_2(b_1, s_2) = \Pi_2(b_1^*, b_{g,2}^*)$.

Proof of Claim 8.3. By Claim 2.1, we have $b_{g,2}^* \in BR_2(b_1)$. Since $s_2 \in BR_2(b_1)$, therefore, we have $\Pi_2(b_1, s_2) = \Pi_2(b_1, b_{g,2}^*)$. But given that any pure strategy in $supp(b_{g,2}^*)$ must reach I, (10) entails that $\Pi_2(b_1, b_{g,2}^*) = \Pi_2(b_1^*, b_{g,2}^*)$. ‖

We now consider the behavioral strategy $\beta_2' \in S_2(\Lambda)$ which is defined as follows:

$$\beta_2'[\emptyset](t) \equiv \begin{cases} 1, & \text{if } t = g \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad \beta_2'[h] = \beta_2^*[h] \text{ for all } h \in H^*(2)\backslash\{\emptyset\}.$$

By Claims 8.2 and 8.3

$$\Pi_2^*(\beta_1^*, \beta_2', \beta_A^*) = \Pi_2(b_1^*, b_{g,2}^*) - G(b_1^*, b_{g,2}^*) - c = \Pi_2(b_1, s_2) - \delta - c. \tag{11}$$

On the other hand, by Lemma 2 and the fact that $s_2 \in BR_2(b_1)$,

$$\Pi_2^*(\beta \,|\, f) = \Pi_2(b_1, b_{f,2}) - F(b_1, s_2) - c = \Pi_2(b_1, s_2) - \delta - c. \tag{12}$$

By using (11), (12) and the hypothesis that $\beta_2^*[\emptyset](\neg D) = 1$, we get $\Pi_2^*(\beta_1^*, \beta_2', \beta_A^*) = \Pi_2^*(\beta \,|\, f) > \Pi_2^{\text{SPE}} = \Pi_2^*(\beta^*)$, which contradicts the sequential rationality of (β^*, μ^*). □

Proof of Theorem 2. Let $(\beta, \mu) \in SE_{\text{w-s}}(\Lambda(\Gamma, \delta, c))$ and $\delta + c < \ell$, where ℓ is chosen small enough that Theorem 1 and Lemmas 7-8 apply. If $\Pi_2^{\text{SPE}} = \max \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, then $\beta_2[\emptyset](\neg D) = 1$ by Lemma 7. If $\Pi_2^{\text{SPE}} < \max \boldsymbol{\Pi}_2^{\text{NE}_1^*}(\Gamma)$, then by Theorem 1, there exists a sequential equilibrium with delegation. Thus by Lemma 8 $\beta_2[\emptyset](\neg D) < 1$, and Lemma 7 completes the proof. □

REFERENCES

Al-Najjar, N. (1995), "A Theory of Forward Induction in Finitely Repeated Games", *Theory and Decision*, **38**, 173-193.

Ben-Porath, E. and Dekel E. (1992), "Signalling Future Actions and the Potential for Sacrifice", *Journal of Economic Theory*, **57**, 36-51.

Brander, J. A. and Lewis, T. R. (1986), "Oligopoly and Financial Structure: The Limited Liability Effect", *American Economic Review*, **76**, 956-970.

Brander, J. A. and Spencer, B. J. (1985), "Export Subsidies and International Market Share Rivalry", *Journal of International Economics*, **18**, 83-100.

Caillaud, B., Jullien, B. and Picard, P. (1995), "Competing Vertical Structures: Precommitment and Renegotiation", *Econometrica*, **63**, 621-646.

Cho, I-K. (1987), "A Refinement of Sequential Equilibrium", *Econometrica*, **55**, 1367-1389.

Cooper, R., DeJong, D., Forsythe, R. and Ross, T. (1993), "Forward Induction in the Battle-of-the-Sexes Games", *American Economic Review*, **83**, 1303-1316.

Dewatripont, M. (1988), "Commitment Through Renegotiation-Proof Contracts with Third Parties", *Review of Economic Studies*, **55**, 377-390.

Fershtman, C. and Gneezy, U. (2001), "Strategic Delegation: An Experiment", *The Rand Journal of Economics*, **32**, 352-368.

Fershtman, C. and Judd, K. (1987), "Equilibrium Incentives in Oligopoly", *American Economic Review*, **77**, 927-940.

Fershtman, C. and Kalai, K. (1997), "Unobserved Delegation", *International Economic Review*, **38**, 763-774.

Fudenberg, D. and Tirole, J. (1991), "Perfect Bayesian Equilibrium and Sequential Equilibrium", *Journal of Economic Theory*, **53**, 236-260.

Hillas, J. (1994), "Sequential Equilibria and Stable Sets of Beliefs", *Journal of Economic Theory*, **64**, 78-103.

Jensen, H. (1997), "Credibility of Optimal Monetary Delegation", *American Economic Review*, **87**, 911-920.

Katz, M. (1991), "Game-Playing Agents: Unobservable Contracts as Precommitments", *Rand Journal of Economics*, **22**, 307-328.

Koçkesen, L. (1999a), "Essays on Preference Formation and Strategic Delegation", Ph.D. Thesis, NYU.

Koçkesen, L. (1999b), "Two-Sided Delegation and Unobservability", (Mimeo.).

Koçkesen, L. and Ok, E. A. (1999), "Strategic Delegation by Unobservable Incentive Contracts", (Mimeo.).

Kohlberg, E. (1990), "Refinement of Nash Equilibrium: The Main Ideas", pp. 3-45 in *Game Theory and Applications* (T. Ichiishi, A. Neyman and Y. Tauman, eds.), San Diego: Academic Press.

Kohlberg, E. and Mertens, J.-F. (1986), "On the Strategic Stability of Equilibria", *Econometrica*, **54**, 1003-1038.

Kreps, D. (1989), "Out-of-equilibrium Beliefs and Out-of-equilibrium Behavior", in F. Hahn (ed.), *The Economics of Missing Markets, Information and Games* (Oxford: Oxford University Press).

Kreps, D. and Wilson, R. (1982), "Sequential Equilibria", *Econometrica*, **50**, 863-894.

McLennan, A. (1985), "Justifiable Beliefs in Sequential Equilibrium", *Econometrica*, **53**, 889-904.

Osborne, M. and Rubinstein, A. (1994) *A Course in Game Theory* (Cambridge: The MIT Press).

Persson, T. and Tabellini, G. (1993), "Designing Institutions for Monetary Stability", *Carnegie-Rochester Conference Series on Public Policy*, **39**, 53-84.

Schelling, T. (1960) *The Strategy of Conflict* (Cambridge: Harvard University Press).

Segendorff, B. (1998), "Delegation and Threat in Bargaining", *Games and Economic Behavior*, **23**, 266-283.

Sklivas, S. (1987), "The Strategic Choice of Managerial Incentives", *Rand Journal of Economics*, **18**, 452-458.

Van Damme, E. (1989), "Stable Equilibria and Forward Induction", *Journal of Economic Theory*, **48**, 476-496.

Vickers, J. (1985), "Delegation and the Theory of the Firm", *Economic Journal*, (supplement) **95**, 138-147.