Equilibrium Selection Through Forward Induction in Cheap Talk Games

Nemanja Antić and Nicola Persico

September 4, 2022

Abstract

This paper provides a refinement that uniquely selects the ex-ante Pareto dominant equilibrium in a cheap talk game, provided one exists. The refinement works by embedding any cheap talk game into a class of two-stage games where: in stage 1 sender and receiver choose their biases at a cost, and in stage 2 the cheap talk game is played. For such games, we show that a forward induction logic can be invoked to select the ex-ante Pareto-dominant equilibrium in the second stage. Games with fixed biases (the conventional cheap talk games) are then treated as limiting cases of this larger class of games.

Nemanja Antić: Managerial Economics and Decision Sciences Department, Kellogg School of Management, Northwestern University, nemanja.antic@kellogg.northwestern.edu;

Nicola Persico: Managerial Economics and Decision Sciences Department, Kellogg School of Management, Northwestern University, and NBER, n-persico@kellogg.northwestern.edu

We would like to thank Joel Sobel for suggesting that we study this problem. We also thank Navin Kartik for helpful comments.

1 Introduction

Cheap talk games are ubiquitous in applied theory. However, cheap talk games have multiple equilibria and this presents a problem for analyzing comparative statics. Applied papers usually restrict attention to the most informative equilibrium, and justify this restriction by invoking the fact that this equilibrium is ex-ante Pareto dominant, i.e., it is the best equilibrium for both sender and receiver. This paper provides a refinement that leverages this intuition and uniquely selects the ex-ante Pareto dominant equilibrium, provided such equilibria exist.[1]

The refinement works by embedding any cheap talk game into a larger space: the class of games where the relative biases of sender and receiver are chosen endogenously before the cheap talk game is played.[2] For such games, we show that a forward induction logic can be invoked to select the ex-ante Pareto-dominant equilibrium in the second stage (Section 5). Games with fixed biases (the conventional cheap talk games) are then treated as limiting cases of this larger class of games.

The model is as follows. Before the cheap talk stage, every agent i obtains a certain quantity of q_i at a cost $c_i(q_i)$. After paying this cost, each agent is assigned the following payoff function in the cheap talk game:

$$U(a, q_i, \omega),$$

where a is the action taken by the receiver, the quantity q_i encodes the heterogeneity across agents, and ω represents the unknown (to the receiver) state of the world. Then the q_i's become publicly observable. Then the sender learns ω and the cheap talk stage takes place.

Given any pair of q_is chosen in the first stage, the second-stage cheap talk game has multiple equilibria. But, in the first stage, agents can use their choice of q_i to "compellingly signal" their expectation that the ex-ante Pareto-dominant equilibrium will be played.

The limiting case with exogenously fixed biases $\overline{q}_i$ can be represented as the limit of sequences of games with cost functions $c_i(\cdot)$ that increasingly penalize any choice $q_i \neq \overline{q}_i$. Our refinement applies to every game in the sequence, and thus selects the Pareto-dominant equilibrium in the limit.

2 Graphical intuition for the result

The result is proved as follows. Fix any first-stage choice $\left(q_i^*, q_{-i}^*\right)$ and assume by contradiction that a Pareto-dominated equilibrium Ω_1 is played in the second stage. We

[1] Crawford and Sobel (1982) provide sufficient conditions for all cheap-talk equilibria to be Pareto-ranked. Therefore, our refinement applies to that class of cheap-talk games. Chen, Kartik, and Sobel (2008) is the only other available refinement which selects the ex-ante Pareto dominant equilibrium, to our knowledge. The pros and cons of that refinement are discussed in Section 7.

[2] Games where the conflict of interest between sender and receiver is determined endogenously prior to the cheap talk phase are of applied interest in their own right. See, for example, Antić and Persico (2020); Argenziano et al., (2016); Austen-Smith, (1994); Deimen and Szalay (2019a, 2019b); Rantakari (2017). These papers, however, do not discuss equilibrium selection.

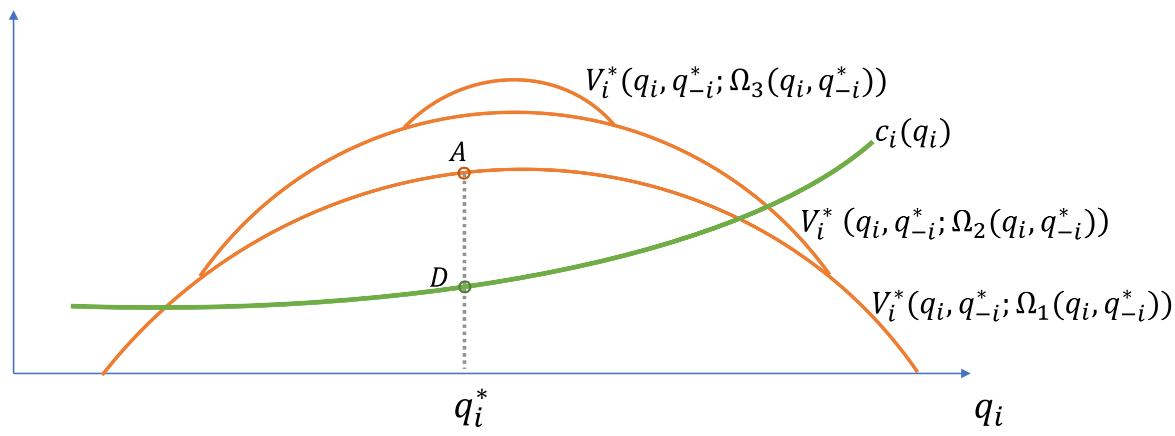


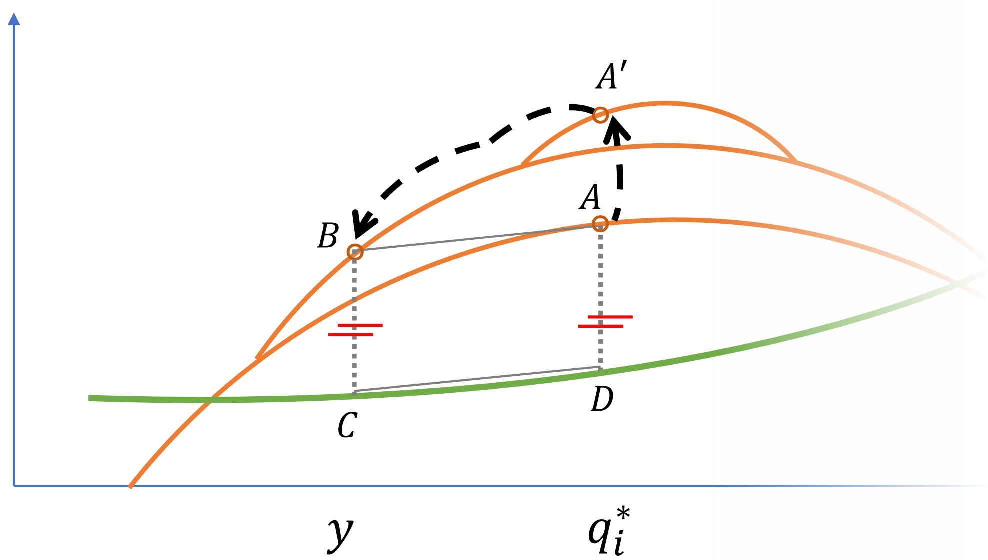
Figure 1: Agent i's best-response problem. The convex green function represents i's cost of choosing a certain level of q_i. Every concave orange dome represents i's ex-ante payoff in the cheap talk game following q_i, if equilibrium Ω is played. Higher domes correspond to equilibria that deliver a higher ex-ante payoff to agent i.

construct a specific deviation $\widetilde{q}_i$ such that, among all second-stage equilibria that are possible following $\left(\widetilde{q}_i, q_{-i}^*\right)$, exactly one of them makes agent i strictly better off, and any other equilibrium makes agent i strictly worse off, relative to i's payoff at $\left(q_i^*, q_{-i}^*, \Omega_1\right)$. The deviation $\widetilde{q}_i$, if it exists, is "compelling" to the other player because it unequivocally coordinates her on the only one among many possible second-stage equilibria that player i "could have wished to achieve" by deviating. If such a deviation exists, the candidate triple $\left(q_i^*, q_{-i}^*, \Omega_1\right)$ is deemed inconsistent with forward induction.

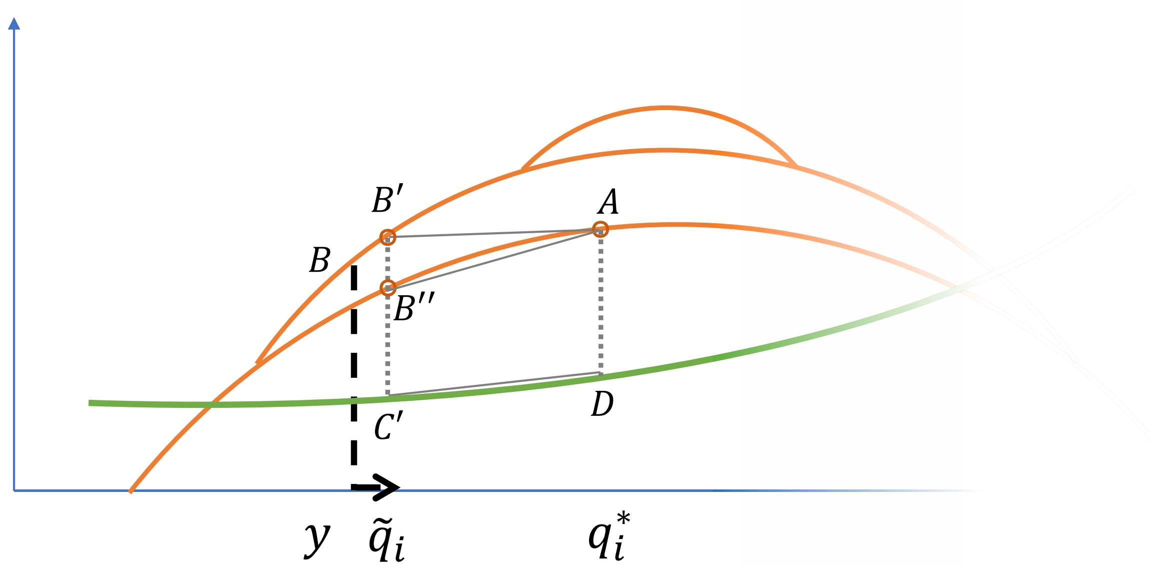
Next we provide a graphical intuition for how the deviation $\widetilde{q}_i$ is constructed. The figures that follow are visually patterned after the quadratic cheap talk game in Crawford and Sobel (1982), but the qualitative properties used in the proof hold much more generally.

For a given q_{-i}^*, Figure 1 illustrates the elements of agent i's best-response problem. The convex green function represents i's cost of choosing a certain level of q_i. Every concave dome represents i's second-stage payoff following q_i, if equilibrium Ω is played. Higher domes correspond to equilibria that deliver a higher ex-ante payoff to agent i. Given q_i^*, for example, Figure 1 indicates that three equilibria exist and that equilibrium Ω_1 is not Pareto-dominant. The segment $\overline{AD}$ represents i's *total* payoff, i.e., the payoff after subtracting the cost c_i, if equilibrium Ω_1 is played following $\left(q_i^*, q_{-i}^*\right)$. We now show that the point A is inconsistent with forward induction.

To construct the deviation $\widetilde{q}_i$, first move up vertically on the graph from point A to point A' (refer to Figure 2 step 1). The segment $\overline{A'D}$ represents i's *total* payoff if the Pareto-dominant equilibrium Ω_3 is played following $\left(q_i^*, q_{-i}^*\right)$. Next, move left along the upper envelope of the orange domes (this corresponds to picking out the best equilibrium following any $q_i < q_i^*$). Initially, that is, for $q_i = q_i^* - \varepsilon$, agent i's total payoff exceeds the total payoff at q_i^* (the latter is represented by the segment $\overline{AD}$). As we continue moving down and to the left along the upper envelope, we will eventually encounter a point $q_i = y$



Step 1: go straight up, then move to the left
along the upper envelope until the point y is reached



Step 2: move back slightly towards q_i^*, and you've found $\widetilde{q}_i$.

Figure 2: This figure illustrates the two-step procedure used to find a "compelling" deviation $\tilde{q}_i$ that agent i can use to coordinate the other agent away from the Pareto-dominated outcome $(q_i^*, q_{-i}^*, \Omega_1)$. The point $\tilde{q}_i$ is such that, among all second-stage equilibria that are possible given $(\tilde{q}_i, q_{-i}^*)$, exactly one of them (Ω_2, corresponding to payoff B') makes agent i strictly better off (this is because $\overline{B'C'} > \overline{AD}$); and all other equilibria make agent i strictly worse off (this is because $\overline{B''C'} < \overline{AD}$).

at which i's total payoff $\overline{BC}$ equals the total payoff $\overline{AD}$. Set $\widetilde{q}_i = y + \varepsilon$: on the graph (Figure 2 step 2), this corresponds to moving back slightly toward q_i^*. The point $\widetilde{q}_i$ is our "compelling" deviation: indeed, among all equilibria that are possible given $\left(\widetilde{q}_i, q_{-i}^*\right)$, exactly one of them (point B') makes agent i strictly better off relative to i's payoff at $\left(q_i^*, q_{-i}^*, \Omega_1\right)$: this is because $\overline{B'C'} > \overline{AD}$; and all other equilibria make agent i strictly worse off (this is because $\overline{B''C'} < \overline{AD}$).

3 General model

This section lays out a more general setting than the one discussed in the introduction, in that the second-stage game need not be a cheap talk game. We study a two-stage sequential game where, in the first stage, two agents indexed by $i = R, S$ simultaneously and independently select q_i at a cost $c_i(q_i)$. We do not assume that $c_i(\cdot)$ is increasing: it could be decreasing or non-monotone. After the first stage, $\mathbf{q} = (q_R, q_S)$ are publicly observed and the second stage is played. The second stage is black-boxed through an index Ω that denotes a second-stage equilibrium given $\mathbf{q} = (q_R, q_S)$. Agent i's payoff in the entire game is:

$$W_i(\mathbf{q}, \Omega) = V_i(q_i, q_{-i}; \Omega) - c_i(q_i), \tag{1}$$

where Ω belongs to $\mathcal{S}(\mathbf{q})$, the set of all equilibria reached at $\mathbf{q}$. The functions $\{V_i\}_{i=R,S}$ capture second-stage payoffs; these functions, and the set $\mathcal{S}(\mathbf{q})$, are independent of the function $c_i(\cdot)$ because, by the second stage, the cost c_i is sunk. The functions $\{V_i, c_i\}_{i=R,S}$ and the set $\mathcal{S}(\mathbf{q})$ are the primitives of the analysis.

A special case is that in which $V_i(q_i, q_{-i}; \Omega)$ represents player i's expected payoff in a cheap talk game where (q_R, q_S) are the players' preference parameters, and Ω identifies which cheap-talk equilibrium is played. This special case will be discussed in Section 5.

When $\mathcal{S}(\mathbf{q})$ has more than one element, checking whether a first-stage choice of q_i is an equilibrium requires one to specify which second-stage equilibria should be used to evaluate a different choice of q_i. With this issue in mind, we define a suitable notion of equilibrium in the sequential game. Note that throughout this section we restrict attention to equilibria $\mathbf{q}^*$ that are in pure strategies.

Definition 1 ***(equilibrium in the sequential game)*** *Let $\mathfrak{s}$ pick, for every $\mathbf{q}$, an element of $\mathcal{S}(\mathbf{q})$. We say that $(\mathbf{q}^*, \mathfrak{s})$ is an equilibrium in the sequential game if, for all i, and all $\widetilde{q}_i$:*

$$W_i(\mathbf{q}^*, \mathfrak{s}(\mathbf{q}^*)) \geq W_i\left(\left(\widetilde{q}_i, q_{-i}^*\right), \mathfrak{s}\left(\widetilde{q}_i, q_{-i}^*\right)\right).$$

This definition parameterizes a (pure-strategy) equilibrium in the sequential game by a selection $\mathfrak{s}$ of second-stage equilibria that are used to evaluate first-stage choices. This corresponds to the usual notion of sequential equilibrium in extensive form games.[3] If, for example, the second stage is a cheap talk game, an example of $\mathfrak{s}$ is: "the cheap talk

[3] Since the actions chosen in the first stage are publically observed, it simplifies notation to just consider the second stage as a stand-alone game with players assigning probability one to histories with the chosen (q_R, q_S).

equilibrium with 2 partition elements if that exists given $\mathbf{q}$, else the babbling equilibrium;" then, first-stage choices would be evaluated by restricting attention to those cheap talk equilibria. In particular, agent i would evaluate her deviations according to the selection $\mathfrak{s}$.

The next definition may be interpreted as an equilibrium selection criterion for second-stage equilibria. It works by restricting "admissible" selections $\mathfrak{s}(\mathbf{q})$ through a criterion in the spirit of forward induction (van Damme 1989).[4]

Definition 2 ***(equilibrium selection through forward induction)*** *An equilibrium* $(\mathbf{q}^*, \mathfrak{s})$ *in the sequential game is consistent with forward induction if no agent* i *and deviation* $\widetilde{q}_i$ *exist, such that exactly one element* $\Omega \in \mathcal{S}\left(\widetilde{q}_i, q^*_{-i}\right)$ *makes agent* i *strictly better off and all other* $\Omega' \in \mathcal{S}\left(\widetilde{q}_i, q^*_{-i}\right)$ *make her strictly worse off.*

This definition starts from a (pure-strategy) equilibrium in the sequential game, and checks for specific deviations $\widetilde{q}_i$ such that, among all second-stage equilibria that are possible following the deviation (not restricted to $\mathfrak{s}\left(\widetilde{q}_i, q^*_{-i}\right)$), exactly one of them makes agent i better off. Such a deviation $\widetilde{q}_i$, if it exists, is "compelling" to the other player because it unequivocally coordinates her on the only one among many possible equilibria in $\mathcal{S}\left(\widetilde{q}_i, q^*_{-i}\right)$ that player i "could have wished to achieve" by deviating.

This criterion for selecting equilibria is conservative in that equilibria are only eliminated by "compelling" deviations which *cannot be misinterpreted by a rational player who believes her opponent to be rational.* Despite being conservative, under some assumptions this criterion will rule out all but one second-stage equilibrium in $\mathcal{S}(\mathbf{q})$.

Definition 3 ***(upper envelope of the payoff correspondence)*** *The upper envelope of* W_i *is the function:*

$$W_i^{\sup}(\mathbf{q}) = \sup_{\Omega \in \mathcal{S}(\mathbf{q})} W_i(\mathbf{q}, \Omega).$$

The upper envelope function selects the upper limit among all the equilibrium payoffs that are possible for agent i given a first-stage choice $\mathbf{q}$. If the set $\mathcal{S}(\mathbf{q})$ of second-stage equilibria is finite, as is the case when $b \neq 0$ in Crawford and Sobel's (1982) quadratic example, the sup operator may be replaced by max.[5] In that example, $W_i^{\sup}$ is attained by the equilibrium with the largest number of partition elements given b.[6]

3.1 Assumptions

The following assumptions are sufficient to prove our main result.

Assumption 1 ***(one-to-oneness)*** *For every* $\mathbf{q}$, $\Omega \neq \Omega'$ *implies* $W_i(\mathbf{q}, \Omega) \neq W_i(\mathbf{q}, \Omega')$.

[4] We acknowledge that Van Damme invokes genericity and finiteness assumptions in his setting, whereas our analysis assumes that $\mathbf{q}$ is selected from a continuum.

[5] Crawford and Sobel (1982) theorem 1 implies that this is true very generally as long as the sender and receiver prefer different actions at all states of the world.

[6] This is true more generally in their setting, provided their condition (M) holds.

This assumption says that, for any given $\mathbf{q}$, second-stage equilibria are one-to-one with payoff levels. In Figure 1 this means that every dome is associated with a distinct cheap talk equilibrium.

Assumption 2 ***(continuous upper envelope)*** *The function* $W_i^{\sup}(\mathbf{q})$ *is continuous for all* i.

Loosely speaking this assumption says that i's best-equilibrium payoff is continuous in $\mathbf{q}$. This assumption says that the upper envelope in Figure 1 is continuous.

Assumption 3 ***(worse option is available)*** *Suppose* $(\mathbf{q}^*, \mathfrak{s})$ *is an equilibrium in the sequential game. Then for every* i *there exists a* $\underline{q}_i$ *such that*

$$W_i^{\sup}\left(\underline{q}_i, q_{-i}^*\right) \leq W_i\left(\mathbf{q}^*, \mathfrak{s}\left(\mathbf{q}^*\right)\right).$$

This assumption says that, for any equilibrium $\mathbf{q}^*$, player i has a "worse option" $\underline{q}_i$ available, at which even the best equilibrium in $\mathcal{S}\left(\underline{q}_i, q_{-i}^*\right)$ gives a (weakly) worse payoff than the equilibrium one. A $\underline{q}_i$ with the required property can always be found if the cost $c_i(\cdot)$ grows fast enough, simply by picking a very large $\underline{q}_i$.

Assumption 4 ***(finite number of second-stage equilibria are pervasive)*** *For any given* q_{-i}*, the set of points* z *such that the set* $\mathcal{S}(z, q_{-i})$ *has finite cardinality, is dense in* $\mathbb{R}$.

This assumption means that for any $\mathbf{q}$ there is an arbitrarily close $\widetilde{\mathbf{q}}$ with a finite number of second-stage equilibria. In Figure 1 this assumption holds because, for *any* given q_i, the number of equilibria is finite (at most three).

4 Results

We say a second-stage equilibrium Ω is Pareto-dominant if all players weakly prefer Ω to any other equilibrium Ω'.

Lemma 1 ***(Pareto-dominant second-stage equilibrium is consistent with forward induction)*** *Consider any equilibrium in the sequential game* $(\mathbf{q}^*, \mathfrak{s})$. *If* $\mathfrak{s}(\mathbf{q})$ *selects a Pareto-dominant second-stage equilibrium in* $\mathcal{S}(\mathbf{q})$ *for all* $\mathbf{q}$*, then* $(\mathbf{q}^*, \mathfrak{s})$ *is consistent with forward induction.*

Proof. See Appendix A. ■

In Crawford and Sobel (1982), the equilibrium with the largest number of partition elements is Pareto-dominant if assumption (M) holds.[7] If $\mathfrak{s}$ selects this equilibrium for

[7] See their Theorems 3 and 5.

every $\mathbf{q}$ then, by the above lemma, any equilibrium $(\mathbf{q}^*, \mathfrak{s})$ in the sequential game is consistent with forward induction.

To see why the above lemma requires $\mathfrak{s}(\mathbf{q})$ to be Pareto-dominant for all $\mathbf{q}$, consider the following simple example in the quadratic setting of Crawford and Sobel (1982). Set both cost functions $\{c_i\} \equiv 0$, and $q_R^* = 2/10, q_S^* = 1/10$. Because $b = q_R^* - q_S^* = 1/10$, the best cheap talk equilibrium has two partition elements. Denote this equilibrium by $\Omega_2 = \mathfrak{s}(q_R^*, q_S^*)$. Suppose $\mathfrak{s}(\mathbf{q})$ picks out the babbling equilibrium for all values of $\mathbf{q}$ such that $q_R - q_S \neq 1/10$ (note that this violates Pareto-dominance). Given this choice of $\mathfrak{s}$, $(\mathbf{q}^*, \mathfrak{s})$ is an equilibrium. However, $(\mathbf{q}^*, \mathfrak{s})$ is not consistent with forward induction. To see this, observe that by increasing q_S slightly above q_S^*, the sender's best-equilibrium payoff exceeds that at $(\mathbf{q}^*, \mathfrak{s})$, but the second-highest equilibrium payoff (babbling is the only other one that exists) does not. Thus, $(\mathbf{q}^*, \mathfrak{s})$ is not consistent with forward induction, but is an equilibrium.

Loosely speaking, this discussion shows that the forward induction requirement "requires" high-payoff outcomes in the second stage on and off the equilibrium path, that is, over the entire domain of the function $\mathfrak{s}(\cdot)$. The next lemma, conversely, says that the only second-stage equilibria that are consistent with forward induction are Pareto-dominant – at least for generic cost functions.

Lemma 2 ***(generically, only Pareto-dominant second-stage equilibria are consistent with forward induction)*** *Suppose Assumptions 1-4 hold. If there is an equilibrium in the sequential game* $(\mathbf{q}^*, \mathfrak{s})$ *such that* $W_i(\mathbf{q}^*, \mathfrak{s}(\mathbf{q}^*)) < W_i^{\sup}(\mathbf{q}^*)$ *for some* i*, then there is a cost function* $\widetilde{c}_i$ *arbitrarily close to* c_i *in the uniform norm, for which:*

1. $(\mathbf{q}^*, \mathfrak{s})$ *remains an equilibrium in the sequential game*
2. $(\mathbf{q}^*, \mathfrak{s})$ *is not consistent with forward induction.*

Proof. See Appendix A. ■

Putting together the two lemmas yields the main result of the paper.

Proposition 1 ***(equilibrium selection of Pareto-optimal equilibria)*** *Suppose that the set* $\{V_i\}_{i=R,S}$, $\mathcal{S}(\mathbf{q})$ *is generated by the cheap talk game in Section 5, and that Assumptions 1-4 hold. Consider any equilibrium* $(\mathbf{q}^*, \mathfrak{s})$.

1. *If* $\mathfrak{s}(\mathbf{q})$ *Pareto-dominates all other cheap talk equilibria for all* $\mathbf{q}$*, then* $(\mathbf{q}^*, \mathfrak{s})$ *is consistent with forward induction.*
2. *If* $\Omega^* = \mathfrak{s}(\mathbf{q}^*)$ *does not Pareto-dominate all other cheap talk equilibria at* $\mathbf{q}^*$*, then there is an* i *and a cost function* $\widetilde{c}_i$ *arbitrarily close to* c_i *in the uniform norm such that* $(\mathbf{q}^*, \mathfrak{s})$ *remains an equilibrium in the sequential game associated with* $\widetilde{c}_i$ *but it is not consistent with forward induction.*

Proposition 1 says the following. Take any equilibrium choice (q_R^*, q_S^*) that is computed under the stipulation that second-stage equilibrium selection is Pareto-dominant "on and off path;" then, this equilibrium is consistent with forward induction (part 1). Furthermore, forward induction requires the second-stage (cheap talk) equilibrium to be Pareto-dominant "on path" (part 2, approximately).

5 Application to cheap talk games

Consider the following special case of the game described in Section 3.

Stage 1 Agents $i = R, S$ simultaneously and independently select a real number q_i at a positive cost $c_i(q_i)$. The q_i's are publicly observed.

Stage 2a Nature selects a state of the world $\omega \in [0, 1]$. The sender privately learns ω and engages in cheap talk with the receiver.

Stage 2b The receiver chooses action $a \in \mathbb{R}$.

Stage 2c An agent i who obtained q experiences utility $U(a, q, \omega)$ in addition to the cost $c_i(q)$.

Note that, in general, the game between S and R is not symmetric because $c_S(\cdot) \neq c_R(\cdot)$. This asymmetry will, in general, cause the agents' equilibrium choices $q_S^* \neq q_R^*$ to not align, resulting in a conflict of interest between sender and receiver in stage 2.

For any given stage-1 choice (q_S, q_R), the cheap talk game described in stage 2 may have multiple equilibria. Let a generic cheap talk equilibrium be indexed by Ω, and let $\mathcal{S}(\mathbf{q})$ denote the set of all cheap talk equilibria given $\mathbf{q} = (q_S, q_R)$. Denote by $V_i(q_i, q_{-i}; \Omega)$ agent i's expected utility U in the cheap talk game with equilibrium $\Omega \in \mathcal{S}(\mathbf{q})$. If the function $W_i(\mathbf{q}, \Omega)$ defined in (1) satisfies Assumptions 1-4, then then the following corollary of Proposition 2 shows that forward induction selects the Pareto-dominant cheap-talk equilibrium, if one exists.

Corollary 1 ***(application to cheap talk games with endogenous conflict of interest)*** *Consider the sequential game described by stages 1 and 2, and let the set* $\{V_i\}_{i=R,S}$, $\mathcal{S}(\mathbf{q})$ *be generated by the game described in stage 2. Suppose Assumptions 1-4 hold. Then, for generic cost functions* c_R *and* c_S*, Proposition 1 shows that forward induction selects the Pareto-dominant cheap-talk equilibrium on and off the equilibrium path, if one exists.*

Corollary 1 requires that Assumptions 1-4 hold. Assumption 1 holds in Crawford and Sobel's (1982) quadratic example this property holds because, for given specification of the bias parameter b, equilibrium payoffs are one-to-one with the number of cutoffs in the equilibrium partition. This is also true more generally when their condition (M) is satisfied.

Assumption 2 holds in Crawford and Sobel's (1982) quadratic example, as shown in Appendix A.2 of Antić and Persico (2020). It also holds in the more general model studied by Crawford and Sobel (1982). Indeed, by their Lemma 3 there is a single equilibrium partition of cardinality N for every b. As b changes, the partitions of this equilibrium deform continuously. Eventually, the equilibrium will cease to exist as partition elements run into each other or into the boundaries of the interval $[0, 1]$. When this happens one of the partition elements vanishes and the cutoffs constitute the equilibrium with cardinality $N - 1$. Since, the receiver's actions are continuous in these cutoffs, so are the players' expected payoffs.

In a cheap talk game, Assumption 3 is satisfied by any $\underline{q}_i$ such that $\mathcal{S}\left(\underline{q}_i, q^*_{-i}\right)$ only contains the babbling equilibrium.[8] In Crawford and Sobel's (1982) quadratic example, the assumption holds if, for any given q_{-i}, agent i's action set includes a q_i such that $|b| = |q_R - q_S| > 1/4$. Assumption 4 holds in Crawford and Sobel's (1982) quadratic example because $b = 0$ is the only value at which the cheap talk game has an infinite number of equilibria. This property also holds under the more general assumptions of Crawford and Sobel's (1982) Theorem 1.

In sum, Assumptions 1-4 are implied by the canonical assumptions in Crawford and Sobel (1982). Moreover, Assumptions 1-4 hold a broader set of cheap talk models including the case of multiplicative bias (e.g., Alonso, 2009; Antić and Persico, 2020).[9]

Finally, we acknowledge that Corollary 1 is only applicable if a Pareto-dominant cheap-talk equilibrium exists. Crawford and Sobel (1982) provide sufficient conditions for all cheap-talk equilibria to be Pareto-ranked. Therefore, our result applies to that class of cheap-talk games.

6 Extension to games with exogenous conflict of interest

The argument derived in this paper can be used to select the best equilibrium in the game with exogenous conflict of interest (i.e., q_i^* is exogenously set equal to $\overline{q}_i$), as follows. The shape of the cost fuctions $c_i(\cdot)$ have been restricted only mildly by our assumptions. In particular, we need not assume $c_i(\cdot)$ is monotone. Consequently, Corollary 1 also applies to cost functions where the cost is prohibitively high except for in a small neighborhood of some exogenous value $\overline{q}_i$. Such cost functions result in equilibrium choices $q_i^* \approx \overline{q}_i$, corresponding to a scenario where the endogeneity in the choice of q_i is almost absent. The limiting case of $q_i^* \equiv \overline{q}_i$ (i.e., fully exogenous preferences in the cheap talk game) can be approximated by a sequence of games where the continuous cost fuctions $c_i(\cdot)$ increasingly penalize any choice $q_i \neq \overline{q}_i$. Intuitively, in games along this sequence the choice of q_i becomes progressively "less endogenous." Since Corollary 1 applies to every game in the sequence it follows that, in the limit game with fully exogenous preferences,

[8] To see this, write:

$$\begin{aligned} W_i(\mathbf{q}^*, \Omega^*) &\geq W_i\left(\left(\underline{q}_i, q^*_{-i}\right), \mathfrak{s}\left(\underline{q}_i, q^*_{-i}\right)\right) \\ &= \sup_{\Omega \in \mathcal{S}\left(\underline{q}_i, q^*_{-i}\right)} W_i\left(\left(\underline{q}_i, q^*_{-i}\right), \Omega\right) = W_i^{\sup}\left(\underline{q}_i, q^*_{-i}\right), \end{aligned}$$

where the inequality holds because $(\mathbf{q}^*, \mathfrak{s})$ is an equilibrium, and the first equality holds because the only element in the set $\mathcal{S}\left(\underline{q}_i, q^*_{-i}\right)$ is the babbling equilibrium.

[9] The canonical Crawford and Sobel (1982) assumptions are: U is bounded and continuous with $U_{11} < 0$, $U_{12} > 0$, $a \in [0,1]$, condition (M) holds and the players disagree about the ideal action for all states. The multiplicative bias case violates Crawford and Sobel's (1982) assumptions because, for exactly one state, the players agree on the ideal action.

the Pareto-dominant equilibrium is the limit of the sequence of forward-induction proof equilibria.

7 Contribution to the literature and conclusions

This paper's contribution is to provide a refinement that uniquely selects the Pareto-dominant equilibrium in a cheap talk game, provided that this equilibrium exists. This is helpful because cheap talk models are used widely and, generally, with an exclusive focus on the Pareto-dominant equilibrium – and yet refinement criteria that uniquely select the Pareto-dominant equilibria are few.

Refinement criteria for signaling games such as Kohlberg and Mertens' (1986) stability and Cho and Kreps' (1987) intuitive criterion, do not have power in cheap talk games because messages are free, unlike costly actions in signaling games. Farrell (1993) introduced qneologism-proofness, which is the first refinement in cheap talk games that does rule out some implausible equilibria, but runs into an existence problem: in many popular cheap talk examples, such as Crawford and Sobel's (1982) quadratic example, no equilibria are neologism-proof. There is a literature that builds on the ideas of neologism proofness. Rabin (1990), Zapater (1997), and Olszewski (2008) face multiplicity: the selection criterion is not stringent enough to select a unique equilibrium. Matthews, Okuno-Fujiwara, and Postlewaite (1991), like Farrell (1993), run into existence problems. Blume and Sobel (1995) propose communication-proof equilibria which, instead of a neologism, allow players an extra opportunity to communicate, however these also need not select the Pareto efficient outcome except in pure common interest games.

The most attractive refinement for cheap talk equilibria is Chen, Kartik, and Sobel's (2008) "no-incentive-to-separate" (NITS) criterion. An equilibrium satisfies NITS if the type-0 sender could not benefit from credibly identifying himself, if he could. This criterion can be microfounded by viewing the cheap talk game as the limit of games with small lying costs. Monotone equilibria in these games converge to the NITS equilibrium in the cheap talk game, as the lying costs converge to zero uniformly. The NITS refinement is attractive because at least one equilibrium always exists that satisfies NITS and, under Crawford-Sobel's "condition M," the only equilibrium that satisfies NITS is the most-informative one. Unlike NITS, our refinement operates at the ex ante stage, i.e., before the sender learns the state of the world. In contrast, in NITS the sender contemplates "causing the equilibrium to switch" after having observed the signal. The two refinements are complementary, in our view: in some applications it may be more natural to assume that there is a cost of lying in the cheap talk game; in other applications, it may be more natural (and even organically desirable) to contemplate the possibility that the agents' biases are determined at an earlier stage of play.

The graphical analysis in Figure 2 is reminiscent of the money-burning logic for equilibrium selection. Roughly speaking, giving a player the option to publicly burn money before a game is played allows that player to "force" others to coordinate on her preferred equilibria in the game without actually having to burn money.[10] Burning money

[10] See Section 4 in van Damme (1989), and Ben-Porath and Dekel (1992).

is somewhat analogous to choosing the compelling deviation $\widetilde{q}_i \neq q_i^*$, except that varying q_i necessarily changes the subsequent cheap talk game, whereas burning money does not because it is a sunk cost. In particular, the choice of q_i affects the equilibrium strategies, payoffs, and even potentially the number of equilibria in the subsequent cheap talk game. This complexity means that the arguments developed in this paper are not straightforward extensions of the money-burning analysis: for example, the compelling deviation $\widetilde{q}_i$ that achieves equilibrium selection in Figure 2 is analogous to *saving* rather than burning money, because $c_i(\widetilde{q}_i) < c_i(q_i^*)$. With this being said, both the money-burning idea and our selection argument are grounded in a forward induction logic and, as such, are conceptually related.

References

[1] Antić, N., and Persico N. (2020). “Cheap Talk with Endogenous Conflict of Interest.” *Econometrica*, 88(6), 2663-2695.

[2] Alonso, R. (2009). “Shared Control and Strategic Communication.” Manuscript, University of Southern California.

[3] Argenziano, R., Severinov, S., and Squintani, F. (2016). “Strategic information acquisition and transmission.” *American Economic Journal: Microeconomics*, 8(3), 119-55.

[4] Austen-Smith, D. (1994). “Strategic Transmission of Costly Information.” *Econometrica*, 62, 955-963.

[5] Ben-Porath, E. and Dekel, E. (1992). “Signaling Future Actions and the Potential for Sacrifice.” *Journal of Economic Theory*, 57, 36-51.

[6] Blume, A. and Sobel, J. (1995). “Communication-proof equilibria in cheap-talk games.” *Journal of Economic Theory*, 65, 359-382.

[7] Chen, Y. , Kartik, N. and Sobel, J. (2008). “Selecting Cheap-Talk Equilibria.” *Econometrica*, 76, 117-136.

[8] Cho, I. K., and Kreps, D. M. (1987). “Signaling games and stable equilibria.” *The Quarterly Journal of Economics*, 102(2), 179-221.

[9] Crawford, V. and Sobel, J. (1982) “Strategic Information Transmission.” *Econometrica*, 50(6), 1431-1451.

[10] Deimen, I., and Szalay, D. (2019a). “Delegated expertise, authority, and communication.” *American Economic Review* 109(4), 1349-1374.

[11] Deimen, I., and Szalay, D. (2019b). “Information and communication in organizations.” *AEA Papers and Proceedings*. 109.

[12] Farrell, J. (1993). “Meaning and credibility in cheap-talk games.” *Games and Economic Behavior*, 5(4), 514-531.

[13] Kohlberg, E. and Mertens, J-F. (1986). “On the Strategic Stability of Equilibria.” *Econometrica*, 54(5), 1003-1037.

[14] Matthews, S. A., Okuno-Fujiwara, M., and Postlewaite, A. (1991). “Refining cheap-talk equilibria.” *Journal of Economic Theory*, 55(2), 247-273.

[15] Olszewski, W. (2006). “Rich language and refinements of cheap-talk equilibria.” *Journal of Economic Theory*, 128(1), 164-186.

[16] Rabin, M. (1990). "Communication between rational agents." *Journal of Economic Theory*, 51(1), 144-170.

[17] Rantakari, H. (2017). "Managerial Influence and Organizational Performance." Manuscript, University of Rochester.

[18] van Damme, Eric (1989) "Stable Equilibria and Forward Induction." *Journal of Economic Theory*, 48(2), 476-496.

[19] Zapater, I. (1997). "Credible proposals in communication games." *Journal of Economic Theory*, 72(1), 173-197.

A Appendix: Equilibrium Selection Through Forward Induction

Proof of Lemma 1

Proof. Because $(\mathbf{q}^*, \mathfrak{s})$ is an equilibrium, for all i we have:

$$W_i(\mathbf{q}^*, \Omega^*) \geq W_i\left(\left(\widetilde{q}_i, q^*_{-i}\right), \mathfrak{s}\left(\widetilde{q}_i, q^*_{-i}\right)\right) \geq W_i\left(\left(\widetilde{q}_i, q^*_{-i}\right), \Omega\right),$$

where $\Omega^* = \mathfrak{s}(\mathbf{q}^*)$ and, by Pareto-dominance of $\mathfrak{s}$, Ω is any element of $\mathcal{S}\left(\widetilde{q}_i, q^*_{-i}\right)$. This means that there is no $\widetilde{q}_i$ and element of $\mathcal{S}\left(\widetilde{q}_i, q^*_{-i}\right)$ that make agent i strictly better off than $(\mathbf{q}^*, \mathfrak{s}(\mathbf{q}^*))$. By definition, this means that $(\mathbf{q}^*, \mathfrak{s})$ is consistent with forward induction. ■

Proof of Lemma 2

Proof. Fix any $\{V_i, c_i\}_{i=R,S}$, $\mathcal{S}(\mathbf{q})$, $\mathfrak{s}(\mathbf{q})$. Take an equilibrium $(\mathbf{q}^*, \mathfrak{s})$ such that for some i:

$$W_i(\mathbf{q}^*, \Omega^*) < W_i^{\sup}(\mathbf{q}^*), \tag{2}$$

where $\Omega^* = \mathfrak{s}(\mathbf{q}^*)$. We now look for a deviation for player i that demonstrates a violation forward induction either for the cost function c_i or for a "nearby" one $\widetilde{c}_i$.

Step 1: identifying the "test neighborhood" $\widetilde{y}$

By Assumption 3, there exists a $\underline{q}_i$ such that

$$W_i^{\sup}\left(\underline{q}_i, q^*_{-i}\right) \leq W_i(\mathbf{q}^*, \Omega^*),$$

which, in conjunction with (2), implies $\underline{q}_i \neq q^*_i$. Assume without loss of generality that $\underline{q}_i > q^*_i$ (the other case is treated symmetrically). By the mean value theorem, which applies because $W_i^{\sup}$ is continuous by Assumption 2, there is at least one $y \geq q^*_i$ such that:

$$W_i^{\sup}\left(y, q^*_{-i}\right) = W_i(\mathbf{q}^*, \Omega^*).$$

Let $\widetilde{y}$ denote the infimum in the set of such y's. This infimum belongs to the set because the set is closed, and thus:

$$W_i^{\sup}\left(\widetilde{y}, q^*_{-i}\right) = W_i(\mathbf{q}^*, \Omega^*), \tag{3}$$

which, in conjunction with (2), implies $\widetilde{y} \neq q^*_i$.

Step 2: identifying a "test deviation" $\widetilde{q}_\eta$ arbitrarily close to $\widetilde{y}$

By Assumption 4, the interval $(q^*_i, \widetilde{y})$ contains an increasing sequence $\{\widetilde{q}_\eta\}_{\eta=1}^{\infty}$ converging to $\widetilde{y}$ such that, for all η, $\mathcal{S}\left(\widetilde{q}_\eta, q^*_{-i}\right)$ is a finite set. Fix any η. Because the set of second-stage equilibria $\mathcal{S}\left(\widetilde{q}_\eta, q^*_{-i}\right)$ is finite, the set $W_i\left(\left(\widetilde{q}_\eta, q^*_{-i}\right), \mathcal{S}\left(\widetilde{q}_\eta, q^*_{-i}\right)\right)$ of associated payoffs has some finite cardinality N_η. By assumption 1 the payoffs in this set may be strictly ordered as follows:

$$w_1(\eta) < w_2(\eta) < \ldots < w_{N\eta}(\eta) = W_i^{\sup}\left(\widetilde{q}_\eta, q^*_{-i}\right). \tag{4}$$

Step 3: constructing the "arbitrarily close" cost function $\widetilde{c}_i(q;\eta)$

Fix any η. Denote

$$\Delta(\eta) = W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) - W_i(\mathbf{q}^*, \Omega^*) > 0, \tag{5}$$

where the inequality follows from the fact that for every $z \in (q_i^*, \widetilde{y})$ we have $W_i^{\sup}\left(z, q_{-i}^*\right) > W_i^{\sup}\left(\widetilde{y}, q_{-i}^*\right) = W_i(\mathbf{q}^*, \Omega^*)$ (both the inequality and the equality follow from the definition of $\widetilde{y}$).

We use the quantity $\Delta(\eta)$ to define an ancillary "upper bound" function that is close to c_i:

$$\overline{c}_i(q;\eta) = c_i(q) + \Delta(\eta) - \frac{1}{2}\min\left[\Delta(\eta), w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right]. \tag{6}$$

The function $\overline{c}_i(\cdot;\eta)$ is greater than $c_i(\cdot)$, but only by a "small amount" less than $\Delta(\eta)$. Now we construct the main object of interest. Let $\widetilde{c}_i(q;\eta)$ be any continuous function such that:

$$\begin{aligned}
\widetilde{c}_i(q;\eta) &\in [c_i(q), \overline{c}_i(q;\eta)] \\
\widetilde{c}_i(\widetilde{q}_\eta;\eta) &= \overline{c}_i(\widetilde{q}_\eta;\eta) \\
\widetilde{c}_i(q_i^*;\eta) &= c_i(q_i^*).
\end{aligned}$$

Step 4: showing that the "test deviation" $\widetilde{q}_\eta$ in conjunction with the cost function $\widetilde{c}_i(q;\eta)$ triggers a violation of forward induction

Fix any η. The payoff function of an agent who is endowed with the cost function $\widetilde{c}_i(q;\eta)$ instead of $c_i(q)$ is denoted by:

$$\widetilde{W}_i(\mathbf{q}, \Omega) = V_i(q_i, q_{-i}; \Omega) - \widetilde{c}_i(q_i;\eta).$$

Denote the ordered elements of the set of payoffs $\left\{\widetilde{W}_i\left(\left(\widetilde{q}_\eta, q_{-i}^*\right), \Omega\right) : \Omega \in S\left(\widetilde{q}_\eta, q_{-i}^*\right)\right\}$ by:

$$\widetilde{w}_1(\eta) < \widetilde{w}_2(\eta) < \ldots < \widetilde{w}_{N\eta}(\eta),$$

with generic element:

$$\widetilde{w}_n(\eta) = w_n(\eta) + c_i(\widetilde{q}_\eta) - \widetilde{c}_i(\widetilde{q}_\eta;\eta). \tag{7}$$

Now note that:

$$\begin{aligned}
\widetilde{w}_{N\eta}(\eta) &= W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) + c_i(\widetilde{q}_\eta) - \widetilde{c}_i(\widetilde{q}_\eta;\eta) \\
&= W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) + c_i(\widetilde{q}_\eta) - \overline{c}_i(\widetilde{q}_\eta;\eta) \\
&= W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) - \Delta(\eta) + \frac{1}{2}\min\left[\Delta(\eta), w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right] \\
&= W_i(\mathbf{q}^*, \Omega^*) + \frac{1}{2}\min\left[\Delta(\eta), w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right] \\
&> W_i(\mathbf{q}^*, \Omega^*) \\
&= \widetilde{W}_i(\mathbf{q}^*, \Omega^*).
\end{aligned}$$

In the above formulas, the first equality comes from the definition of $\widetilde{w}_n(\eta)$ (eq. 7), after using (4) to substitute for $w_{N\eta}(\eta)$. The next equality holds because $\widetilde{c}_i(\widetilde{q}_\eta;\eta) = \overline{c}_i(\widetilde{q}_\eta;\eta)$ by construction. The equality in line 3 results from substituting for $\overline{c}_i$ using (6), and the next equality follows from substituting for $\Delta(\eta)$ using (5). The strict inequality holds because both arguments of the min operator are strictly positive. The final equality holds because $\widetilde{c}_i(q_i^*;\eta) = c_i(q_i^*)$ by construction.

Proceeding as in the previous paragraph:

$$
\begin{aligned}
\widetilde{w}_{N\eta-1}(\eta) &= w_{N\eta-1}(\eta) + c_i(\widetilde{q}_\eta) - \widetilde{c}_i(\widetilde{q}_\eta;\eta) \\
&= w_{N\eta-1}(\eta) - \Delta(\eta) + \frac{1}{2}\min\left[\Delta(\eta), w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right] \\
&\leq w_{N\eta-1}(\eta) - \Delta(\eta) + \frac{1}{2}\left[w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right] \\
&< w_{N\eta-1}(\eta) - \Delta(\eta) + \left[w_{N\eta}(\eta) - w_{N\eta-1}(\eta)\right] \\
&= W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) - \Delta(\eta) \\
&= W_i(\mathbf{q}^*, \Omega^*) \\
&= \widetilde{W}_i(\mathbf{q}^*, \Omega^*).
\end{aligned}
$$

In the above formulas, the equality in line 5 uses (4) to substitute for $w_{N\eta}(\eta)$.

In sum, we have shown that

$$\widetilde{w}_{N\eta-1}(\eta) < \widetilde{W}_i(\mathbf{q}^*, \Omega^*) < \widetilde{w}_{N\eta}(\eta). \tag{8}$$

Step 5: wrap up

We have shown that, for every η:

Under $\widetilde{c}_i(q;\eta)$, the pair $(\mathbf{q}^*, \mathfrak{s})$ remains an equilibrium in the sequential game.

This is true because q_i^* remains a best response to q_{-i}^* (this follows because, by construction $\widetilde{c}_i(q_i^*;\eta) = c_i(q_i^*)$ and $\widetilde{c}_i(q;\eta) \geq c_i(q)$).

Under $\widetilde{c}_i(q;\eta)$, the pair $(\mathbf{q}^*, \mathfrak{s})$ is not consistent with forward induction.

This is true because equation (8). Note that while this equation is a statement about *payoffs*, also has implications for *second-stage equilibria* in light of Assumption 1. Indeed, equation (8) shows that the deviation $\widetilde{q}_\eta$ is such that exactly one element of $\mathcal{S}\left(\widetilde{q}_\eta, q_{-i}^*\right)$ makes agent i strictly better off and all others make her strictly worse off. Thus, the deviation $\widetilde{q}_\eta$ is used to show that the equilibrium $(\mathbf{q}^*, \mathfrak{s})$ is not consistent with forward induction.

It remains to show that $\widetilde{c}_i(q;\eta)$ and $c_i(q)$ can be made arbitrarily close by an opportune choice of η. To verify this, observe that, for all q,

$$0 \leq \widetilde{c}_i(q;\eta) - c_i(q) < \Delta(\eta).$$

Substitute (3) into (5) to get:

$$\Delta(\eta) = W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) - W_i^{\sup}\left(\widetilde{y}, q_{-i}^*\right).$$

By construction, the sequence $\{\widetilde{q}_\eta\}_{\eta=1}^{\infty}$ converges to $\widetilde{y}$, and by continuity of $W_i^{\sup}$ (Assumption 2) we get:

$$\lim_{\eta\to\infty} \Delta(\eta) = \lim_{\eta\to\infty} W_i^{\sup}\left(\widetilde{q}_\eta, q_{-i}^*\right) - W_i^{\sup}\left(\widetilde{y}, q_{-i}^*\right) = 0.$$

This shows that a suitable choice of η makes $\Delta(\eta)$ the uniform distance between $\widetilde{c}_i(q;\eta)$ and $c_i(q)$ arbitrarily small. ■