Iterated Exclusion of Implausible Types in Signaling Games

Francesc Dilmé

December 6, 2023

Abstract

Cho and Kreps (1987) proposed a series of criteria for selecting equilibria in signaling games. Their procedure for applying each criterion was to identify all implausible sender types associated with a given off-path message, then look for sequential equilibria assigning probability zero to every implausible type. This paper proposes a different selection procedure for each given selection criterion: the *iterated exclusion* of implausible types. We show that this procedure has more selection power, is easier to implement, is independent of the exclusion order, and selects equilibrium outcomes with higher internal consistency. We prove that sequentially stable outcomes, which exist in all finite signaling games, pass the iterated exclusion procedure. Moreover, we show how the procedure can be used to establish whether or not a given outcome is sequentially stable.

Keywords: Equilibrium selection, signaling, game theory, sequential stability.

JEL Codes: C72, D82, D83.

University of Bonn. `fdilme@uni-bonn.de`. This work was funded by a grant from the European Research Council (ERC 949465).

1 Introduction

Since the seminal paper of Spence (1973), the study of signaling games has been central to information economics, with applications in settings such as job markets, insurance markets, and bargaining. While signaling games are often simple, their study is complicated by their high equilibrium multiplicity, which makes it necessary to narrow predictions through equilibrium selection. In the literature, this is most commonly done using the selection criteria introduced by Cho and Kreps (1987)—namely, from weakest to strongest, the Intuitive Criterion (IC), D1, D2, and Never a Weak Best Response (NWBR).[1] These "classical" selection criteria have played a major role in the analysis of signaling games and are often used to determine theoretical predictions in applications.[2]

Cho and Kreps proposed to apply each selection criterion as follows. Fix a sequential outcome ω and an off-path message m; then identify all sender types t for whom sending m is implausible according to the criterion. (For example, the IC says that if the type-t sender obtains a strictly lower payoff from sending m than from behaving as specified by the outcome ω, *independently* of the sender's response to m, then m is implausible for type t.) After identifying all implausible pairs (t, m), check whether there is a sequential equilibrium with outcome ω in which, after observing m, the receiver assigns positive probability only to plausible types. The outcome ω passes the selection criterion if and only if such a sequential equilibrium exists.

While this procedure ensures a certain degree of internal consistency, it fails to exclude some outcomes that are fragile to repeated application of the logic described. Indeed, an outcome passing a given criterion may be supported only by sequential equilibria in which, after observing an off-path message m, the receiver assigns positive probability to types who cannot benefit from choosing m for any best response of the receiver to beliefs assigning probability zero to the types pruned out by the criterion. It would be natural for such types to be pruned out as well. To ensure full consistency, one would have to prune types iteratively until no further types could be pruned out, and then require that supporting sequential equilibria assign probability zero to all pruned types.

In this paper we introduce a selection method based on the iterative application of the classical selection criteria, with two goals in mind: (1) to select "fully consistent" equilibrium outcomes (i.e., outcomes immune to the above criticism), and (2) to do so in a simple and intuitive manner. For each

[1]Banks and Sobel (1987) proposed different selection methods, called Divinity and Universal Divinity, which are iterative but significantly more difficult to use. Here we focus on generalizing the simpler methodology of Cho and Kreps (1987).

[2]Although the criteria of Cho and Kreps (1987) have proved to be the most popular, other criteria have also been proposed; see the literature review.

criterion $X \in \{\mathrm{IC}, \mathrm{D1}, \mathrm{D2}, \mathrm{NWBR}\}$, we define a procedure called *iterated exclusion through* X (IEX), which works as follows. Fix a sequential outcome ω and an off-path message m. In each step, there is a standing set of pairs that were excluded in the previous steps. The criterion X is used to exclude the next type–message pair. (For example, if $X = \mathrm{IC}$, then the pair (t, m) is excluded if the type-t sender obtains a strictly lower payoff from sending m than from behaving as specified by ω, for all of the receiver's best responses to beliefs assigning probability zero to the previously excluded types.) The outcome ω passes IEX if, for every m, after the exclusion of all implausible pairs, there is a sequential equilibrium with outcome ω in which, after observing m, the receiver assigns positive probability only to non-excluded types.[3] We show that if, at some step in the iteration, there is no sequential equilibrium with outcome ω in which, after observing m, the receiver assigns probability zero to all currently excluded types, then ω fails IEX. Hence it may not be necessary to run IEX exhaustively to assess whether an outcome passes.

We find that IEX has desirable properties for all criteria X. First, it is stronger than the classical selection procedure: If an outcome passes IEX, then it passes the procedure of Cho and Kreps (1987) for the criterion X, while the converse may not be true. Second, the order of exclusion is irrelevant to the procedure's output. This implies that one can exclude multiple pairs in each step, but one need not exclude all currently implausible types before moving to the next step. Third, because it is iterative and the results are independent of the order of exclusion, IEX is not subject to certain criticisms applied to the classical procedure, in which two pairs may simultaneously exclude each other—that is, one pair may be excluded owing to the presence of the second, while the second is excluded owing to the presence of the first. Finally, each IEX is easier to apply than the classical procedure X, because, as the set of non-excluded types shrinks after each step, the set of the receiver's responses to beliefs assigning positive probability only to non-excluded types shrinks as well, making it easier to verify whether each remaining type should be excluded.

In the second part of the paper, we study the relationship between iterated exclusion and sequentially stable outcomes (Dilmé, 2023b). We first prove that any sequentially stable outcome passes every IEX. Hence, since every signaling game has a sequentially stable outcome, there is an outcome passing each IEX. Furthermore, if there is a unique outcome passing IEX for some X, then such an outcome is both sequentially stable and strategically stable (Kohlberg and Mertens, 1986). We provide two results characterizing the sequential stability of outcomes passing IEX.

[3] If all types are excluded at some step, ω passes IEX if there is a sequential equilibrium where the receiver assigns probability one to the type pruned out in the last iteration.

Overall, we see iterated exclusion as a convenient and effective tool for equilibrium selection in signaling games, encompassing and strengthening the classical selection criteria. We hope it will allow a more consistent equilibrium selection across applications.

1.1 Literature review

Ever since their introduction by Spence (1973), signaling games have been acknowledged to have very high equilibrium multiplicity, which limits their ability to make predictions in applications. Indeed, most signaling games feature not only a plethora of Nash equilibria (Nash, 1950), but also large sets of equilibria satisfying various forms of sequential rationality, such as subgame perfect equilibria (Selten, 1965), trembling-hand perfect equilibria (Selten, 1975), or sequential equilibria (Kreps and Wilson, 1982).

The dominant approach to resolving this problem is via equilibrium selection criteria tailored to signaling games.[4] Most applications use the selection criteria proposed by Cho and Kreps (1987), which are based on ruling out certain equilibria by imposing formal restrictions on off-path beliefs.[5] Our work takes these criteria a step further by applying them iteratively; this leads to a stronger selection method, which we also argue is easier to use.

Some of the equilibrium refinements in the literature are aimed at eliminating outcomes that are not strategically stable (Kohlberg and Mertens, 1986; see the discussions in Cho and Kreps, 1987, and Banks and Sobel, 1987). In this paper, however, we focus on linking our iterated exclusion procedure to sequential stability (Dilmé, 2023b) rather than to strategic stability. This is because sequential stability and iterated exclusion both apply to outcomes (whereas strategic stability applies to connected sets of Nash equilibria), and sequentially stable outcomes admit a simple characterization in signaling games. For generic payoffs, sequential stability coincides with strategic stability.

The rest of the paper is organized as follows. In Section 2, we define our signaling game and related notation. In Section 3 we briefly recall the four classical selection criteria, then define the iterated exclusion process and describe how to use it. Section 4 describes the connection between iterated exclusion and sequential stability. Finally, Appendix A contains the proofs of the results.

[4] This approach is also widespread in the cheap-talk literature, since communication games are often degenerate signaling games. Works introducing selection criteria for communication games include Farrell (1993), Chen et al. (2008), Kartik (2009), and Dilmé (2023a).

[5] See McLennan (1985) and Cho (1987) for a similar approach in extensive-form games. Other examples of selection criteria in signaling games can be found in Banks and Sobel (1987), Mailath et al. (1993), and Carlsson and Dasgupta (1997).

2 Signaling games

In this section we define our signaling game, which coincides with that of Cho and Kreps (1987), and provide some notation. We will keep the signaling game fixed throughout the paper.

A signaling game proceeds as follows. First, nature chooses a type $t \in T$ with distribution $\mu_0 \in \Delta(T)$ having full support. After observing t, the sender chooses a message $m \in M_t$. Finally, having observed the message but not the type, the receiver chooses a response $r \in R_m$. We assume that T, $M := \cup_{t \in T} M_t$, and $R := \cup_{m \in M} R_m$ are finite. We let $T_m \subset T$ be the set of types for which the sender can send message m. We let $u_t(m,r)$ and $u_{\mathrm{r}}(t,m,r)$ denote the payoffs of the sender and the receiver, respectively. For a given outcome ω (i.e., a distribution over terminal histories), we let $u_t(\omega)$ denote the sender's payoff conditional on the realized type being t.

A strategy σ for the sender assigns, to each $t \in T$, some mixed choice of messages $\sigma_t \in \Delta(M_t)$. A strategy ρ for the receiver assigns, to each $m \in M$, a mixed reply $\rho_m \in \Delta(R_m) =: \mathcal{R}_m$.

Sequential equilibria

Following Cho and Kreps (1987), we base our equilibrium selection process on the concept of sequential equilibrium (Kreps and Wilson, 1982), which we now define.

A *belief system* μ assigns a posterior $\mu_m \in \Delta(T_m)$ to each $m \in M$. For a given belief system μ and message m, we let $\mathrm{BR}_m(\mu_m) \subset \mathcal{R}_m$ be the set of (mixed) best responses of the receiver to m and belief μ_m. We let $\mathrm{BR}_m := \cup_{\mu_m \in \Delta(T_m)} \mathrm{BR}_m(\mu_m)$ be the set of all best responses of the receiver, across all beliefs. Finally, for each $T' \subset T_m$, we let $\mathrm{BR}_m(T') := \cup_{\mu_m \in \Delta(T')} \mathrm{BR}_m(\mu_m)$ be the set of all best responses of the receiver to beliefs with support in T'.

An *assessment* is a pair comprising a strategy profile and a belief system. A *sequential equilibrium* is an assessment (σ, ρ, μ) such that (i) the sender best-responds to the receiver's strategy (i.e., σ is optimal given ρ), (ii) the receiver is sequentially rational given the posterior (i.e., $\rho_m \in \mathrm{BR}_m(\mu_m)$ for all m), and (iii) the assessment is *consistent* (i.e., each μ_m is the limit of the type distribution conditional on m through a sequence of fully mixed sender's strategies). The following result states that, in signaling games, consistency is equivalent to the on-path Bayes rule.

Lemma 2.1. *An assessment is consistent if and only if it satisfies the on-path Bayes rule.*

Lemma 2.1 establishes that, in signaling games, consistency does not dictate any restriction on off-path beliefs. The implication is that focusing on sequential equilibria (as opposed to Nash equilibria, for example) imposes a minimal plausibility requirement: that, off path, the receiver best-responds to some beliefs about the sender's types. For future convenience, we let SE_ω denote the set of all sequential equilibria with a given outcome ω, and for all $m \in M$ and $T' \subset T_m$, we let

$$\mathrm{SE}_{\omega,m}(T') := \left\{ (\sigma, \rho, \mu) \in \mathrm{SE}_\omega \,\middle|\, \mu_m(t) = 0 \ \forall t \in T_m \setminus T' \right\} .$$

We say that ω is a *sequential outcome* if there is a sequential equilibrium with outcome ω (i.e., if $\mathrm{SE}_\omega \neq \emptyset$).

3 Selection criteria and iterated exclusion

In this section, we first present the classical selection criteria of Cho and Kreps (1987), then introduce our procedure for iterated exclusion through each criterion, and finally provide some examples.

3.1 The Intuitive Criterion, D1, D2, and Never a Weak Best Response

Here we briefly recall the classical criteria of Cho and Kreps (1987): the Intuitive Criterion (IC), D1, D2, and Never a Weak Best Response (NWBR). We fix an outcome ω and an off-path message m (note that, unlike Cho and Kreps, we do not require that ω is first verified to be a sequential outcome before applying our method). For notational convenience, we assume that $T_m = T$, and to make the analysis non-trivial, we assume that $|T_m| \geq 2$. For each $T' \subset T$ with $T' \neq \emptyset$ and each $t \in T$, we define

$$\begin{aligned} D_t(T') &:= \left\{ \rho_m \in \mathrm{BR}_m(T') \,\middle|\, u_t(m, \rho_m) > u_t(\omega) \right\} \text{ and} \\ D_t^0(T') &:= \left\{ \rho_m \in \mathrm{BR}_m(T') \,\middle|\, u_t(m, \rho_m) = u_t(\omega) \right\} . \end{aligned}$$

Note that, unlike Cho and Kreps, we allow these sets to depend on a set of types; this will be useful when we analyze the iterated exclusion procedures introduced in Section 3.2.

We now define four statements about $T' \subset T$ and $t \in T$, one for each criterion:

$$\begin{aligned} &\mathrm{IC}_t(T') \text{ holds if} && D_t(T') \cup D_t^0(T') = \emptyset , \\ &\mathrm{D1}_t(T') \text{ holds if} && D_t(T') \cup D_t^0(T') \subset D_{t'}(T') \text{ for some type } t' \in T , \\ &\mathrm{D2}_t(T') \text{ holds if} && D_t(T') \cup D_t^0(T') \subset \cup_{t' \in T \setminus \{t\}} D_{t'}(T') , \\ &\mathrm{NWBR}_t(T') \text{ holds if} && D_t(T') \cup D_t^0(T') \subset \cup_{t' \in T} D_{t'}(T') . \end{aligned}$$

As explained in the introduction, Cho and Kreps (1987) propose the following selection method for each criterion $X \in \{\mathrm{IC}, \mathrm{D1}, \mathrm{D2}, \mathrm{NWBR}\}$. In Step 1, all implausible types (according to criterion X) are pruned out.[6] That is, Step 1 is to identify the set of plausible types $\hat{T}$, which is the complement of the set of implausible types

$$T \setminus \hat{T} := \{\, t \in T \mid X_t(T) \text{ holds} \,\} \, .$$

In Step 2, one checks whether there is a sequential equilibrium assigning probability zero to all pruned types (i.e., those in $T \setminus \hat{T}$): If yes, ω passes the criterion X; if not, it fails it.

Observations

We now make several observations about the classical selection criteria.

First, note that our version of the IC is slightly different from that of Cho and Kreps (1987). We have made this alteration for consistency, that is, to put the IC on the same footing as the other selection criteria. Cho and Kreps, in their selection procedure for the IC, do not use Step 2 as described above; instead, they say that an outcome fails IC if there is a type who strictly profits from deviating to m for all receiver best responses to beliefs assigning zero probability to pruned types. Our version has slightly greater selection power, as fewer outcomes pass it.

Second, note that if $D_t(T) \cup D_t^0(T) = \emptyset$ for all $t \in T$, then each of the criteria IC, D1, D2, and NWBR prunes out all types; thus, no sequential equilibrium assigns probability zero to all pruned types. Still, it seems clear that the outcome should not be deemed implausible in this eventuality. Cho and Sobel (1990) address this issue by saying that an outcome survives criterion D1 if one can find a corresponding sequential equilibrium where, "for all off-the-equilibrium-path signals m, $\mu_m(t) = 0$ whenever $D_t(T) \cup D_t^0(T) \subset D_{t'}(T)$ holds for some t' such that $D_{t'}(T) \neq \emptyset$" (p. 385). Similar fixes can be used for the other criteria. As we shall see, our iterated exclusion procedure will also sometimes prune out all types.

Third, note that our definition of $\mathrm{NWBR}_t(T')$ is not the same as the one in Cho and Kreps (1987).

[6] Given that we undertake our analysis for a fixed ω and m, with our procedure iteratively reducing the set of types for whom sending m is plausible, we will speak of pruning or excluding types t, rather than type–message pairs (t, m) as in Cho and Kreps (1987).

However, the formulations are equivalent: It is clear that for all T' and t,

$$D_t(T') \cup D_t^0(T') \subset \cup_{t' \in T} D_{t'}(T') \quad \text{if and only if} \quad D_t^0(T') \subset \cup_{t' \in T \setminus \{t\}} D_{t'}(T') \,. \tag{3.1}$$

We believe that our definition of $\text{NWBR}_t(T')$ elucidates the connection between the different criteria, and it also makes obvious their relative strength: Visual inspection shows that, for all T' and t,

$$\text{IC}_t(T') \text{ holds} \;\Rightarrow\; \text{D1}_t(T') \text{ holds} \;\Rightarrow\; \text{D2}_t(T') \text{ holds} \;\Rightarrow\; \text{NWBR}_t(T') \text{ holds} \,. \tag{3.2}$$

That is, NWBR prunes out more types than D2, which prunes out more types than D1, which prunes out more types than IC (Grossman and Perry, 1986, Banks and Sobel, 1987, and Cho and Kreps, 1987, show that the reverse implications do not hold in general). Often, however, the pruning out of types with weaker criteria is technically less involved.

Finally, we present a generalization of a result that is stated but not formally proven by Cho and Kreps (1987), and that gives the NWBR criterion its name.

Lemma 3.1. *For all t and T',* $\text{NWBR}_t(T')$ *holds if and only if there is no sequential equilibrium in* $\text{SE}_{\omega,m}(T')$ *where m is a weak best response for t.*

The intuition for Lemma 3.1 is that, if there is a sequential equilibrium $(\sigma, \rho, \mu) \in \text{SE}_{\omega,m}(T')$ where $\rho_m \in D_t^0(T')$ (so m is a best response for t), then it must be that $\rho_m \notin \cup_{t' \neq t} D_{t'}(T')$, since otherwise there would be a type $t' \in T$ who would strictly benefit from choosing m. Conversely, if $D_t^0(T') \not\subset \cup_{t' \neq t} D_{t'}(T')$, one can construct a sequential equilibrium with outcome ω where m is a best response for t. Note that our result applies to any $T' \subset T$, not just $T' = T$.

3.2 Iterated exclusion through a selection criterion

The motivation for using the classical selection criteria is that they ensure some degree of internal consistency: An outcome passes a given criterion if it is supported by a sequential equilibrium assigning probability zero, after each off-path message, to all types who cannot benefit from deviating to that message. However, the sequential equilibrium may fail a further consistency requirement: It may assign positive probability to some types who cannot benefit from deviating under any plausible beliefs of the receiver (i.e., any beliefs assigning a positive probability only to non-pruned types); see Example 3.1.

In this section we present an alternative selection method, based on the iterative application of the classical selection criteria, which will be shown to ensure full consistency. For each $X \in \{\text{IC}, \text{D1}, \text{D2}, \text{NWBR}\}$, we define a procedure called *iterated exclusion through criterion X* ("IEX" for short).[7] Intuitively, for a given outcome and message, the IEX process iteratively excludes types that are implausible according to criterion X, given the current set of excluded types. The formal definition is as follows:

> **Iterated exclusion through criterion X (IEX).** Initialize $T_1 = T$. Then, in each step $n \geq 1$, if $X_t(T_n)$ holds for some $t \in T_n$, set $T_{n+1} := T_n \setminus \{t\}$. Otherwise, set $T_{n+1} := T_n$.

This iterative process is aimed at identifying whether ω is the outcome of some sequential equilibrium with reasonable receiver beliefs. Roughly, it works as follows. In the first iteration, we rule out some type t_1 that is implausible given m, according to the selection criterion X; thus, we set $T_1 = T \setminus \{t_1\}$. In the second iteration, we exclude a type t_2 that is implausible, according to X, given that t_1 has already been excluded; thus, $T_2 = T_1 \setminus \{t_1\}$. (For example, if $X = \text{IC}$, then t_2 is excluded if, for all of the receiver's best responses to beliefs assigning probability zero to t_1, type t_2 strictly loses from deviating to m.) In the third iteration, we exclude a type t_3 that is implausible given that t_1 and t_2 have both been excluded, and so on.

Note that IEX converges in at most $|T|$ steps. Note also that there are potentially different "implementations" of IEX, since, if there are two types that could be excluded at a given step, the procedure does not specify which one to exclude first. We now establish that once IEX has converged (i.e., when no more types can be excluded), the set of excluded types is independent of the order of exclusion.

Proposition 3.1. *For all X, the set of types excluded by applying* IEX *is independent of the order of exclusion.*

From now on, we denote by $T^X_{\omega,m}$ the set of types not excluded after running any implementation of IEX. The fact that $T^X_{\omega,m}$ is independent of the order of exclusion is important for several reasons. First, it implies that unlike some procedures, such as the iterated elimination of weakly dominated strategies, IEX produces the same results regardless of how it is implemented. Second, it allows one to exclude a set of types, instead of a single type, in each round; this is especially useful in large games.[8] The IEX procedure is therefore stronger than applying criterion X à la Cho and Kreps (1987) (see Examples 3.1 and 3.2). It is also easier to use: Since T_n shrinks as types get excluded,

[7] If, for example, $X = \text{D1}$, we say "iterated exclusion through D1" (or IED1), and if $X = \text{NWBR}$, we abuse language and say "iterated exclusion of never-a-weak-best-response" (or IENWBR).

[8] The results of IEX are also independent of the order in which sets of types are excluded, but this does not follow directly from Proposition 3.1. Rather, it follows from an intermediate result in the proof (Lemma A.1) establishing that if a given type t can be excluded in a given step n of an implementation of IEX, then t can be excluded in all the steps that follow n.

verifying whether $X_t(T_n)$ holds becomes easier at each step. Finally, Proposition 3.1 permits us to define a *canonical implementation* of IEX in which, at each step n, all types for whom $X_t(T_n)$ holds are excluded.[9] Under the canonical implementation of IEX, the first step of our procedure coincides with Step 1 in the selection method of Cho and Kreps for criterion X.

Having defined IEX as a procedure for excluding types, we can now state what it means for an outcome to pass or fail IEX:

Definition 3.1. For all X, we say that ω *passes* IEX if, for all off-path m, either (i) $\mathrm{SE}_{\omega,m}(T^X_{\omega,m}) \neq \emptyset$, or (ii) $T^X_{\omega,m} = \emptyset$ and, letting $\hat{t}$ be the last type excluded through some implementation of IEX, we have $\mathrm{SE}_{\omega,m}(\{\hat{t}\}) \neq \emptyset$. Otherwise, we say that ω *fails* IEX.

Let us shed some light on this definition. If $T^X_{\omega,m} \neq \emptyset$ and $\mathrm{SE}_{\omega,m}(T^X_{\omega,m}) \neq \emptyset$, then ω is supported by a sequential equilibrium that assigns a positive probability only to types not excluded by criterion X, even after the criterion has been applied iteratively as many times as possible; hence IEX does not rule out ω (for m). If $T^X_{\omega,m} \neq \emptyset$ and $\mathrm{SE}_{\omega,m}(T^X_{\omega,m}) = \emptyset$, then there is no such sequential equilibrium supporting ω; hence IEX rules out ω. Finally, if $T^X_{\omega,m} = \emptyset$, then ω is said to pass if, under *some* implementation of IEX, there was a sequential equilibrium supporting ω before the last type was excluded. We note that if $T^X_{\omega,m} = \emptyset$, then, by the proof Proposition 3.1, the property "$\mathrm{SE}_{\omega,m}(\{\hat{t}\}) \neq \emptyset$ for the last excluded type $\hat{t}$" is independent of the order of exclusion in IEX; in other words, if ω passes (under condition (ii) of Definition 3.1) for *some* implementation of IEX, then it passes for *all* implementations of IEX. (The proof of this result is highly non-trivial, because different implementations of IEX may entail the exclusion of types corresponding to significantly different sets of sequential equilibria in each step.) Condition (ii) of Definition 3.1 can thus be interpreted as saying that ω is supported by a sequential equilibrium assigning a positive probability only to non-excluded types at every stage of every implementation of IEX. In fact, it is easy to see that ω passes IEX and $T^X_{\omega,m} = \emptyset$ only if there is a sequential equilibrium where all types strictly lose from sending m (but note that the existence of one such equilibrium is not sufficient for ω to pass IEX for m).

Our motivation for using IEX instead of X is to find *internally consistent* sequential equilibria supporting a given outcome. Consider, for example, the criterion $X = \mathrm{IC}$, which requires that the receiver assigns probability zero to the types who cannot possibly gain from choosing m if the receiver chooses a best response to some belief. Clearly, if m is chosen, then the receiver's beliefs should assign

[9] A disadvantage of the canonical implementation of IENWBR is that, in each step, one needs to exclude *all* types for whom m is never a weak best response; this can be difficult in some games. Thanks to Proposition 3.1, however, one can use a different implementation and obtain the same results.

probability zero to the initially pruned types, and the receiver should best-respond accordingly. If there is a type that cannot possibly gain from choosing m under these circumstances, it makes sense to exclude this type as well. It is easy to see that $T^X_{\omega,m}$ is the biggest set T' satisfying the property "$t \in T'$ if and only if type t can (weakly) benefit from choosing m for some best response of the receiver to some belief with support in T'." Hence, if $T^X_{\omega,m} \neq \emptyset$, there is an internally consistent sequential equilibrium with outcome ω when $\mathrm{SE}_{\omega,m}(T^X_{\omega,m}) \neq \emptyset$, while there is no such equilibrium when $\mathrm{SE}_{\omega,m}(T^X_{\omega,m}) = \emptyset$.[10]

Remark 3.1 (Divinity and universal divinity). Banks and Sobel (1987) define two selection criteria, called *divinity* and *universal divinity*, which, similarly to our IEX, consist in the iterated exclusion of types. However, these criteria are difficult to apply, which is why the criteria of Cho and Kreps (1987) have been more widely used. As Cho and Kreps point out, divinity and universal divinity are roughly based on requiring D2 and that, if $D_t(T) \cup D^0_t(T)$ is nonempty, then $\mu_m(t) \geq \mu_0(t)$. Banks and Sobel prove that divine equilibrium outcomes exist for all games, and universally divine equilibrium outcomes exist for generic signaling games.

Early assessment that ω fails IEX

The following result establishes that IEX need not be applied exhaustively until there are no more excludable types: If, in some step, there is no sequential equilibrium with outcome ω where the receiver assigns probability zero to the currently excluded types, then ω fails IEX.

Proposition 3.2. *For all X, if $T_n \neq \emptyset$ and $\mathrm{SE}_{\omega,m}(T_n) = \emptyset$ at some step of an implementation of* IEX*, then ω fails* IEX.

Proposition 3.2 gives researchers a convenient way to assess the plausibility of an outcome without having to apply IEX exhaustively. Broadly, of course, IEX is rendered easier to use by the fact that one need not check that $\mathrm{SE}_{\omega,m}(T_n) = \emptyset$ at each step n. However, if at some step the researcher intuits that ω will fail IEX, then they can check whether $\mathrm{SE}_{\omega,m}(T_n) = \emptyset$; if so, it follows that ω fails.

Why exclusion (and not elimination)?

Some equilibrium selection procedures studied in the literature consist in recursively eliminating actions from the game at each step; that is, when an action is considered "implausible", the analysis

[10] If $T^X_{\omega,m} = \emptyset$, then ω passes IEX if there is a sequential equilibrium that assigns a positive probability to a single type and such that no type wants to deviate; otherwise, ω fails IEX.

moves to a simpler game where this action and the histories following it are not present. The advantage of such procedures is that the plausibility of equilibria is easier to evaluate in the simpler game. This is the case, for example, in the iterated elimination of weakly/strictly dominated strategies (Bernheim, 1984, and Pearce, 1984), and in elimination through NWBR (Kohlberg and Mertens, 1986, and Dilmé, 2023b). We now briefly discuss the advantage of "excluding" types (i.e., keeping them in the game, but looking for the receiver's best responses to beliefs assigning them probability zero) instead of "eliminating" them (removing them from the game).

In each step n of our recursive procedure, given the current set of non-excluded types T_n, two sets determine procedure continuation: the set of the receiver's best responses $\mathrm{BR}_m(T_n)$ (which determines $D_t(T_n)$ and $D_t^0(T_n)$ for each t), and the set of sequential equilibria $\mathrm{SE}_{\omega,m}(T_n)$ (which determines the plausibility of the outcome). When types are excluded (and not eliminated), both sets decrease as n increases; that is, $\mathrm{BR}_m(T_{n+1}) \subset \mathrm{BR}_m(T_n)$ and $\mathrm{SE}_{\omega,m}(T_{n+1}) \subset \mathrm{SE}_{\omega,m}(T_n)$. On the other hand, if types are eliminated, then the set of receiver best responses still decreases at each step, but the set of sequential equilibria may not.[11] In that case, an outcome may pass a criterion thanks to a sequential equilibrium that is supported through strategies that are not "reasonable" in the original game (see Example 3.4), or the assessment of a criterion may depend on the order of elimination (see Example 3.5).

3.3 Examples

Example 3.1. We illustrate that IEIC has strictly more selection power than IC, by analyzing a simple example. Consider the game in Table 1, with $x=2$ and $y=-1$.[12] Let $\omega_{m'}$ indicate the outcome where all types choose m' for sure. Note that, after m, $1\circ r_1$ is a best response to $1\circ t_1$, and that $u_{t_1}(m, 1\circ r_1)=1>u_{t_1}(\omega_{m'})$, so $\mathrm{IC}_{t_1}(T)$ does not hold. Similarly, $1\circ r_3$ is a best response to $1\circ t_3$ and we have $u_{t_2}(m, 1\circ r_3)=1>u_{t_2}(\omega_{m'})$, so $\mathrm{IC}_{t_2}(T)$ does not hold. Since, for type t_3, sending m is strictly dominated by sending m', $\mathrm{IC}_{t_3}(T)$ holds, so IC only prunes out t_3. Note that $\omega_{m'}$ is the outcome of a sequential equilibrium where $\rho_m=1\circ r_2$ and $\mu_m=1\circ t_2$; therefore, $\omega_{m'}$ passes IC. But this equilibrium is not intuitive: it is sustained by the belief that m is sent by t_2, who strictly loses from sending m for any best response of the receiver to beliefs assigning probability 0 to the pruned type t_3.

[11] While there are fewer receiver best responses to beliefs in $\Delta(T_{n+1})$ than to beliefs in $\Delta(T_n)$ (because $T_{n+1} \subset T_n$), a sequential equilibrium needs to satisfy fewer incentive constraints in the game where only types in T_{n+1} can send m than in the game where only types in T_n can send m.

[12] The notation we use is common in the literature: Table 1 describes a game containing three types, $T=\{t_1,t_2,t_3\}$, and two messages, $M=\{m,m'\}$. The receiver has one reply to m' (that is, $R_{m'}=\{r_0\}$) and three replies to m (that is, $R_m=\{r_1,r_2,r_3\}$). For each combination of type and response, the table gives the sender's payoff (first number) and the receiver's payoff (second number).

m'	r_0
t_1	0,0
t_2	0,0
t_3	0,0

m	r_1	r_2	r_3
t_1	$1,1$	$-2,-x$	$-1,-1$
t_2	$-2,-1$	y,x	$2,-2$
t_3	$-1,0$	$-1,1$	$-1,2$

Table 1: Payoffs from the games in Examples 3.1 and and 3.5.

Next, continuing the IEIC process, observe that if the receiver's belief assigns probability zero to t_3, then r_3 is dominated by both r_1 and r_2, and when the receiver plays r_3 with probability zero, type t_2 obtains a strictly lower payoff from playing m than from playing m'. Hence, $\mathrm{IC}_{t_2}(\{t_1,t_2\})$ holds. Also, $\mathrm{IC}_{t_1}(\{t_1,t_2\})$ fails to hold, for the same reason as before. Therefore, t_2 is excluded in the second round. It is clear that there is no sequential equilibrium where the sender assigns probability zero to $\{t_2,t_3\}$; hence, $\omega_{m'}$ fails IEIC.[13]

Example 3.2. We now illustrate that IENWBR has strictly more selection power than NWBR. Consider a signaling game where $T=\{1,...,10\}$, $M=\{m,m'\}$, $R_{m'}=\{0\}$, and $R_m=\{1,...,10\}$. All types obtain 0 from choosing message m' (necessarily followed by the receiver choosing $r=0$). Assume that if the sender sends m, then the receiver wants to match the sender's type, while each type t benefits from choosing m only if the sender choice is close to $t+1$. More concretely, we assume

$$u_{\mathrm{r}}(m,r)=\mathbb{I}_{r=t} \text{ and } u_t(m,r)=2-(t+1-r)^2 .$$

In this game, the outcome $\omega_{m'}$ where all types choose m' passes NWBR. Indeed, note that type $t=10$ can be pruned out: if she is indifferent between choosing m or m', then type $t=9$ is strictly willing to choose m. Note then that there are sequential equilibria (σ,ρ,μ) with outcome $\omega_{m'}$ where $\mu_m(10)=0$ (for example, there is one where $\mu_m(t)=\rho(r)=1/9$ for all $t,r\in\{1,...,9\}$). Hence, $\omega_{m'}$ passes NWBR.

Once the receiver considers type 10 implausible, the second iteration pf IENWBR becomes very similar to the first. The previous reasoning can now be used to prune out type 9: If ρ is a best response to a posterior μ_m with $\mu_m(10)=0$, then $\rho(10)=0$, and so we have that if such ρ makes type 9 indifferent between m and m', then type 8 is strictly willing to choose m. The same logic is applied to prune out types 8, 7, ..., until only type 1 is left (type 1 cannot be further pruned out). Is then there

[13] Like Cho and Kreps (1987), we do not wish to rely too much on "speeches" to defend our selection criteria. Still, for this example, one could think of a deviation by the sender as implicitly making the following speech: "I am sending message m to convince you I am type t_1. Indeed, I would never wish to send m if I were t_3, as this would give me a lower payoff than sending m', independently of what you do. Hence, since r_3 is a dominated strategy for you once you rule out t_3, you should deduce that I am not type t_2 either, because if I were, both r_1 and r_2 would give me a payoff strictly lower than if I had sent m'."

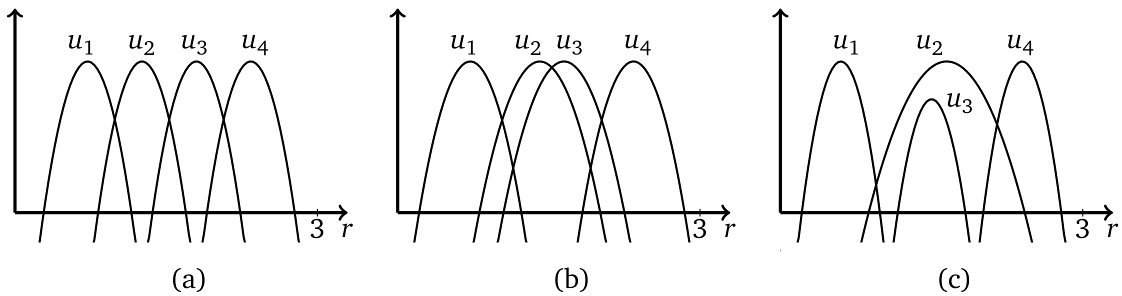

Figure 1: Figure corresponding to Example 3.3. Each u_t is short for $u_t(m,\cdot)$.

a sequential equilibrium where $\mu_m(1)=1$? The answer is no: the receiver best responds to $\mu_m(1)=1$ by choosing $r=1$, but then type 1 has a strict incentive to choose m. Hence, $\omega_{m'}$ fails IENWBR.

Example 3.3. This example illustrates how the irrelevance of the exclusion order established in Proposition 3.1 overcomes some criticism directed at D2 (and sometomes at NWBR). Consider a signaling game where $T=\{0,1,2,3\}$, $M=\{m,m'\}$, $R_{m'}=\{0\}$, and R_m is a finite but large set of actions belonging to $[0,3]$. The receiver wants to match the type, $u_{\mathrm{r}}(t,m,r)=-(t-r)^2$. As a result, if R_m is large and spread enough within $[0,3]$, any best response to a posterior will consist of at most two close-by responses, those closest to the expected type. Each type obtains 0 from choosing m' (necessarily followed by the receiver choosing $r=0$). We consider three possibilities for the payoffs of the sender after sending m, depicted in the panels depicted in Figure 1, respectively (for visual clarity, we depict them as functions from $[0,3]$ to $\mathbb{R}$). We focus on an outcome $\omega_{m'}$ where all types send m'.

Because all types benefit from some best response in all panels, no type can be pruned out using IC. Additionally, only in panel (c) can a type be pruned out using D1 (type 3). In panel (b), both type 2 and type 3 can be pruned out using D2; yet, if one of them was not present, the other type could not be pruned out. This fact draws some criticism by Cho and Kreps (1987), who say, "One can construct examples in which [...] one type is eliminated by virtue of several others, one of which is simultaneously eliminated because the first type is not yet eliminated. That is, each helps to eliminate the other. We are, in consequence, happier with tests built up out of D1 than with those built up out of D2" (p. 208). Nonetheless, note that if type 2 is excluded in the first iteration of IED2, then type 3 can be excluded in the second iteration: Since type 2 belongs to $T\backslash\{3\}$ independently of whether it is pruned out or not, and since

$$D_3(T')\cup D_3^0(T')\subset\cup_{t'\in T\backslash\{3\}}D_{t'}(T')$$

m'	r_0
t_1	0,0
t_2	0,0

m	r_1	r_2	r_3
t_1	$-1,3$	$1,2$	$-1,0$
t_2	$1,0$	$1,2$	$-2,3$

Table 2: Payoffs from the game in Example 3.4.

holds for $T'=\{0,1,2,3\}$, this condition also holds for $T'=\{0,1,3\}$. Therefore, through the iterated exclusion of types using IED2, it becomes natural that if two types help exclude the other in a given iteration, each of these types will eventually be excluded at some iteration, even when the other is already excluded. Indeed, while excluded types are not entertained by the receiver in each iteration, they are still used to assess the implausibility of the remaining types, because their incentive not to deviate has hold in the resulting sequential equilibria supporting the outcome. Finally, note that types 2 and 3 can be pruned out using NWBR in all panels, and that in panel (a), they help exclude the other. Furthermore, in panel (c), if type 2 was not present, type 3 could not be eliminated using NWBR. Still, even if type 2 is excluded in the first iteration of IENWBR, type 3 can be eliminated in the second iteration.

Example 3.4. This example illustrates the advantages of exluding types instead of eliminating them. Consider the example in Table 2, which corresponds to the game in Figure IV in Cho and Kreps (1987). Let $\omega_{m'}$ indicate the outcome where both types choose m' for sure. Cho and Kreps note that, while no type can be pruned out under $\omega_{m'}$ using D1, t_2 can be pruned out using NWBR.[14] Since there is no sequential equilibrium where the receiver assigns probability 1 to type t_1 after m, $\omega_{m'}$ fails NWBR (as well as IENWBR). However, if t_2 is eliminated (instead of excluded), the game where t_2 is not present has a sequential equilibrium. In this equilibrium, the receiver plays r_1 after m, which is reasonable within the smaller game but which induces t_2 to deviate in the original game.

Cho and Kreps find the fact that $\omega_{m'}$ fails NWBR—because t_2 strictly benefits from deviating in any best response of the sender excluding t_2—"downright unintuitive". We, however, see NWBR (or, more generally, IEX for any X) as a systematic method for evaluating the internal consistency of an equilibrium outcome. In the game in Table 2, $\omega_{m'}$ is inconsistent according to NWBR because the sender should rule out type t_2 after m, but there is no sequential equilibrium where she does so.

Example 3.5. In this example, we again illustrate the convenience of excluding instead of eliminating types, now showing that the assessment of the procedure may depend on the order of elimination if

[14]Note that the set of sequential equilibria is $(\mu(t_1),\sigma(r_1),\sigma(r_2))\in((2/3,1]\times\{0\}\times\{0\})\cup(\{2/3\}\times\{0\}\times[0,1/2])$. In all sequential equilibria, r_1 is strictly dominated, and m is dominated for type t_2.

one does not verify in each step that the set of sequential equilibria with support on non-eliminated types is nonempty. Consider again the game in Table 1, now with $x=-2$ and $y=-1$. It is easy to see that $\omega_{m'}$ (the outcome where all types choose m' for sure) has a sequential equilibrium where $\rho_m = \frac{1}{2}\circ r_1 + \frac{1}{2}\circ r_3$ and $\mu_m = \frac{1}{2}\circ t_1 + \frac{1}{2}\circ t_3$. We note that, under this equilibrium, m is a weak best response for both t_1 and t_2, and that (for the same reason as in Example 3.1) $\text{NWBR}_{t_3}(T)$ holds.[15]

Now consider the game where t_3 cannot send m. Then m is a strictly dominant action for t_1, while it is a strictly dominated action for t_2. Hence, in this smaller game, $D^0_{t_1}(\{t_1, t_2\}) = D^0_{t_2}(\{t_1, t_2\}) = \emptyset$ (recall (3.1)), so both $\text{NWBR}_{t_1}(T)$ and $\text{NWBR}_{t_2}(T)$ hold. If t_1 is eliminated first, then the game where only t_2 can send m has a sequential equilibrium, so it "passes" iterated elimination through NWBR. If instead t_2 is eliminated first, then the game where only t_1 can send m has no sequential equilibrium, so it "fails" iterated elimination through NWBR. Note that if we use exclusion instead of elimination, then the set of sequential equilibria before the last excluded type is empty under both orders of exclusion; that is, $\text{SE}_{\omega,m}(\{t_1\}) = \text{SE}_{\omega,m}(\{t_2\}) = \emptyset$. Hence the assessment of IENWBR is independent of the order of exclusion, as established in the proof of Proposition 3.1.

4 Sequential stability in signaling games

The concept of strategically stable sets of equilibria (Kohlberg and Mertens, 1986) has been very influential in the literature aiming to refine the set of Nash equilibria, because of its many desirable properties (such as existence, forward induction, iterated dominance, invariance, and robustness to small trembles). Cho and Kreps (1987) argue that part of the value of their selection criteria stems from their partial characterization of strategically stable equilibria, which are often difficult to obtain.

In an earlier paper (Dilmé, 2023b), we defined the related concept of sequentially stable outcomes and showed that they have similar properties to strategically stable equilibria, while being easier to use. In this section, we briefly review strategic stability and sequential stability, then use IEX to partially characterize each of them.

4.1 Strategically and sequentially stable outcomes

Here we review the concepts of strategic stability and sequential stability, and we recall a simple characterization of the latter for signaling games.

[15] It is easy to see that, under IENWBR, t_2 is excluded next, and since there is no sequential equilibrium where the receiver assigns probability zero to $\{t_2, t_3\}$, it follows that $\omega_{m'}$ fails IENWBR.

Strategic stability was introduced by Kohlberg and Mertens (1986) as an equilibrium concept for extensive-form games. Their idea was to perturb the reduced normal form of the game through trembles and look for behavior that was stable to *any* perturbation. They defined a *strategically stable set of equilibria* as a set of Nash equilibria of the reduced normal form of the game with the property that, for any vanishing sequence of fully mixed trembles, there is a corresponding sequence of Nash equilibria converging to some element of the stable set. They then defined a *strategically stable component* as a minimal (by set inclusion) stable set of equilibria. They showed that, in games with generic payoffs, all elements of a stable component have the same outcome; such outcomes are called *strategically stable outcomes*.

Dilmé (2023b) introduced the concept of a sequentially stable outcome, which also applies to extensive-form games. An outcome ω is defined as *sequentially stable* if, for all vanishing sequences of fully mixed behavioral trembles, there is a corresponding sequence of perfect ε_n-equilibria with outcomes converging to ω, for some $\varepsilon_n \to 0$.[16] Dilmé (2023b) shows that all games have a sequentially stable outcome. For signaling games, strategically stable outcomes are sequentially stable, and if there is a unique sequentially stable outcome, then it is the unique strategically stable outcome. For signaling games with generic payoffs, the two concepts coincide.

We now present a simple characterization of sequential stability for signaling games.

Proposition 4.1 (Proposition 5.1 in Dilmé, 2023b)**.** *An outcome ω is sequentially stable if and only if it is sequential and, for any off-path $m \in M$ and $\mu_m \in \Delta(T)$, there are some $\alpha \in [0,1]$, $\mu'_m \in \Delta(T)$, and $\rho_m \in \mathrm{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu'_m)$ such that $u_t(m,\rho_m) \le u_t(\omega)$ for all t, and such that if $\alpha \neq 1$, then $u_t(m,\rho_m) = u_t(\omega)$ for all t with $\mu'_m(t) > 0$.*

Remarkably, Proposition 4.1 implies that in a signaling game, one can check sequential stability message by message, even though sequential stability requires stability to trembles across all messages.[17] The characterization makes it much easier to analyze sequential stability in signaling games, as our discussion below illustrates (see, in particular, the proofs of Propositions 4.2–4.4).

[16] In a perfect ε_n-equilibria, all players ε_n-optimize at all information sets *conditionally* on their being reached.

[17] The possibility of analyzing each off-path message independently greatly simplifies the arguments for sequential stability (in comparison to those for strategic stability). The situation is similar to that of sequential equilibria, which have a message-by-message characterization (recall Lemma 2.1), in comparison to trembling-hand perfect equilibria (Selten, 1975).

4.2 Sequential stability and IEX

In this section, we state several results which establish connections between sequential stability and IEX.

Sequentially stable outcomes pass IEX

The first result in this section establishes that sequential stability is stronger than IEX.

Proposition 4.2. *For all X, sequentially stable outcomes pass IEX.*

Given that sequentially stable outcomes always exist, Proposition 4.2 implies that every signaling game has an outcome that passes IEX. This is a valuable result, because often it is easy to prove that an outcome fails a selection criterion, but considerably harder to prove that an outcome passes. Hence, Proposition 4.2 allows us to prove that a given outcome passes IEX by checking that all other candidates (e.g., sequential outcomes) fail. It also lets us prove that a given outcome is *not* sequentially stable by proving it does not pass IEX. This is helpful because the characterization of sequential stability in Proposition 4.1 is sometimes difficult to apply. Furthermore, the fact that every signaling game has an outcome passing IEX permits us to compare predictions across applications. Finally, given that each IEX is stronger than X, it is more likely that there is only one outcome passing IEX. If this is the case, such an outcome is not only guaranteed to be the only sequentially stable outcome of the signaling game but also its only strategically stable outcome.

The proof of Proposition 4.2 illustrates how Proposition 4.1 helps in proving results. The key insight is that, since $\mathrm{SE}_{\omega,m}(T_{n+1}) \subset \mathrm{SE}_{\omega,m}(T_n)$, if a type t is excluded at some step n, then $\mathrm{NWBR}_t(T_{n'})$ holds for all $n' \geq n$. Assume then that ω is sequentially stable and that $\mathrm{SE}_{\omega,m}(T_n) \neq \emptyset$ (which holds for at least $n=1$). Let t_n be the type excluded in step n; then $X_{t_n}(T_n)$ holds, so $\mathrm{NWBR}_{t_n}(T_n)$ holds as well (recall the expression (3.2)). Take some $\mu_m \in \Delta(T_{n+1})$. Because m is never a best response for an equilibrium in $\mathrm{SE}_{\omega,m}(T_{n+1})$ for any of the types excluded in the previous steps, it must be that μ''_m (defined in Proposition 4.1) assigns probability zero to $T \setminus T_{n+1}$, which implies that $\mathrm{SE}_{\omega,m}(T_{n+1}) \neq \emptyset$.[18]

Example 4.1. This example shows that the converse to Proposition 4.2 is not true: There may be outcomes that are *not* sequentially stable, but that pass IENWBR (and therefore pass IEX for all X). To see this, consider Table 3, which corresponds to Figure 3 in Banks and Sobel (1987). Let $\omega_{m'}$

[18]From Proposition 4.1 it is easy to see that if ω is sequentially stable, then for any $\mu_m \in \Delta(T)$, there are some $\alpha \in [0,1]$, $\mu'_m \in \Delta(T)$, and $(\sigma, \rho, \mu'') \in \mathrm{SE}_\omega$ such that $\mu''_m = \alpha \mu_m + (1-\alpha)\mu'_m$, and such that if $\alpha \neq 1$ and $\mu'_m(t) > 0$, then m is a weak best response for t under (σ, ρ, μ'').

m'	r_0
t_1	0,0
t_2	0,0

m	r_1	r_2	r_3	r_4
t_1	$-1,3$	$-1,2$	$1,0$	$-1,-2$
t_2	$-1,-2$	$1,0$	$1,2$	$-2,3$

Table 3: Payoffs from the game in Example 4.1.

indicate the outcome where both types choose m' for sure. It is easy to see that, for each type, there is a sequential equilibrium with outcome $\omega_{m'}$ where m is a weak best response. In the first of these equilibria, denoted by $(\sigma,\hat{\rho},\hat{\mu})$, we have $\hat{\rho}_m:=\frac{1}{2}\circ r_3+\frac{1}{2}\circ r_4$ and $\hat{\mu}_m:=\frac{1}{3}\circ t_1+\frac{2}{3}\circ t_2$ (in this case, $u_{t_1}(m,\hat{\rho}_m)=0$ and $u_{t_2}(m,\hat{\rho}_m)=-1/2$). In the second, denoted by $(\sigma,\check{\rho},\check{\mu})$, we have $\check{\rho}_m:=\frac{1}{2}\circ r_1+\frac{1}{2}\circ r_2$ and $\check{\mu}_m:=\frac{2}{3}\circ t_1+\frac{1}{3}\circ t_2$ (in this case, $u_{t_1}(m,\check{\rho}_m)=-1$ and $u_{t_2}(m,\check{\rho}_m)=0$). Hence, $\omega_{m'}$ passes IENWBR.

Banks and Sobel show that an equilibrium with outcome $\omega_{m'}$ is *not* part of a strategically stable set of equilibria. We use an argument similar to (but simpler than) theirs to show that $\omega_{m'}$ is not sequentially stable. For the sake of contradiction, we assume that $\omega_{m'}$ is sequentially stable, then apply Proposition 4.1. Fix the belief $\mu_m:=0.51\circ t_1+0.49\circ t_2$. We have that $\mathrm{BR}_m(\mu_m)=\{1\circ r_2\}$, and note that $u_{t_2}(m,1\circ r_2)=2$. Then, since m is never simultaneously chosen by both types for any best response of the receiver, there are only two possibilities. The first is that $\mu'_m=1\circ t_1$ and $\mu''_m=\check{\mu}_m$ (where μ'_m and μ''_m are as in the statement of Proposition 4.1), but this cannot hold, because $u_{t_1}(m,\check{\rho}_m)<0$. The second possibility is that $\mu'_m=1\circ t_2$ and $\mu''_m=\hat{\mu}_m$, but this cannot hold either, because $u_{t_2}(m,\hat{\rho}_m)<0$.

Ruling out that an outcome is sequentially stable

As shown by Example 4.1, although all sequentially stable outcomes pass IEX, an outcome may pass IEX yet fail to be sequentially stable. The next proposition shows that even if an outcome passes IEX, the process of iterated exclusion can be used to prove that it is not sequentially stable.

Proposition 4.3. *Let $T_n\neq\emptyset$ be the set of non-excluded types remaining after (not necessarily exhaustively) applying IEX, and assume $\mathrm{SE}_{\omega,m}(T_n)\neq\emptyset$. Then, for any $\mu_m\in\Delta(T_n)$, if there is some $\mu'_m\in\Delta(T)$ satisfying the conditions in Proposition 4.1, then $\mu'_m\in\Delta(T_n)$.*

Proposition 4.3 indicates that to prove an outcome is *not* sequentially stable, one can apply IEX to exclude some types for some message $m\in M$. Then, fixing a posterior on the set of non-excluded types T_n, one only needs to check that there is no posterior in $\Delta(T_n)$ satisfying the property in Proposition 4.1; that is, one need not check all of $\Delta(T)$.

Establishing that an outcome is sequentially stable

We now show that IEX can be used not only to rule out sequential stability, but also to prove it.

Proposition 4.4. *Fix a sequential outcome ω and an off-path message m. Assume that there are $T' \subset T_m$ and $\rho_m \in \mathrm{BR}_m(T')$ such that $u_t(m,\rho_m)=u_t(\omega)$ for all $t \in T'$, and $u_t(m,\rho_m) \leq u_t(\omega)$ for all $t \notin T'$. Then (i) ω is sequentially stable if and only if it is sequentially stable in the game where m is not available, and (ii) $T' \subset T^X_{\omega,m}$ for all X.*

Proposition 4.4 provides a tool for proving that a given message is *not* an impediment to the sequential stability of a given outcome.[19] The first part of the proposition is a corollary of Proposition 4.1. Indeed, if T' and ρ_m satisfy the hypotheses of Proposition 4.4, then we can fulfill the condition in Proposition 4.1 by setting $\alpha := 1$ and choosing $\mu'_m \in \Delta(T')$ so that $\rho_m \in \mathrm{BR}_m(\mu')$. The second part of Proposition 4.4 establishes that if some T' with the properties in the statement exists, it is a subset of $T^X_{\omega,m}$. Hence, the iterated exclusion process makes it easier for a researcher to identify a suitable T' in order to prove that ω passes IEX for m.

Example 4.2. We illustrate the use of Proposition 4.4 using the game in Table 1, now with $x=2$ and $y=4$. Consider the outcome $\omega_{m'}$ where the sender chooses m' for sure. It is clear that since m is strictly dominated for type t_3, this type can be eliminated in the first round of IEX for all X. Then, it is not difficult to see that there is a sequential equilibrium where both t_1 and t_2 obtain payoff 0 when playing m (in this equilibrium, $\rho_m = \frac{2}{3} \circ r_1 + \frac{1}{3} \circ r_2$ and $\mu_m = \frac{1}{2} \circ t_1 + \frac{1}{2} \circ t_2$), so $\hat{T} := \{t_1, t_2\}$ satisfies the conditions in Proposition 4.4. Therefore, $\omega_{m'}$ is sequentially stable.

[19]Note that, from Proposition 4.1, if ω is sequentially stable and m is off-path, then ω is sequentially stable in the game where m is not available.

A Proofs

Proof of Lemma 2.1

Proof. Let (σ,ρ,μ) be an assessment. If it is consistent, it is clear that it satisfies the on-path Bayes rule. Assume then that (σ,ρ,μ) satisfies the on-path Bayes rule, and we will show it is consistent. Consider the sequence of fully-mixed sender's strategies $(\sigma_n)_{n\geq\overline{n}}$ defined by

$$\sigma_{t,n}(m):=\begin{cases}\mu_m(t)/\mu_0(t)\,n^{-1} & \text{if } \sigma_t(m)=0 \text{ and } \mu_m(t)>0,\\ n^{-2} & \text{if } \sigma_t(m)=0 \text{ and } \mu_m(t)=0,\\ K_{t,n}\,\sigma_t(m) & \text{if } \sigma_t(m)>0,\end{cases}$$

for all $t\in T$ and $m\in M_t$, where $K_{t,n}$ is such that $\sum_{m\in M_t}\sigma_{t,n}(m)=1$, and where $\overline{n}$ is the minimum such that $K_{t,n}>0$ for all $n\geq\overline{n}$. Consider also the sequence of fully-mixed receiver's strategies $(\rho_n)_{n\geq\hat{n}}$ defined by

$$\rho_{m,n}(r):=\begin{cases}n^{-1} & \text{if } \rho_m(r)=0,\\ \hat{K}_{m,n}\,\rho_m(r) & \text{if } \rho_m(r)>0,\end{cases}$$

for all $m\in M$ and $r\in R_m$, where $\hat{K}_{m,n}$ is such that $\sum_{r\in R_m}\rho_{m,n}(r)=1$, and where $\hat{n}$ is the minimum such that $\hat{K}_{m,n}>0$ for all $n\geq\hat{n}$. It is easy to see that, because (σ,ρ,μ) satisfies on-path Bayes rule, the sequence of strategy profiles $(\sigma_n,\rho_n)_{n\geq\max\{\overline{n},\hat{n}\}}$ supports (σ,ρ,μ), hence (σ,ρ,μ) is consistent. □

Proof of Lemma 3.1

Proof. **"Only if" part.** Assume $\mathrm{NWBR}_t(T')$ holds. Assume also, for the sake of contradiction, that m is a best response for t in some sequential equilibrium $(\sigma,\rho,\mu)\in\mathrm{SE}_{\omega,m}(T')$. This implies that $\rho_m\in\mathrm{BR}_m(T')$ is such that $u_t(m,\rho_m)=u_t(\omega)$. By the assumption that $\mathrm{NWBR}_t(T')$ holds, there is a type $t'\in T$ such that $u_{t'}(m,\rho_m)>u_{t'}(\omega)$, but this contradicts that (σ,ρ,μ) is a sequential equilibrium and that m is off path.

"If" part. Assume there is *no* sequential equilibrium in $\mathrm{SE}_{\omega,m}(T')$ where m is a best response for t. For the sake of contradiction, assume that $\mathrm{NWBR}_t(T')$ does not hold. This implies that there is some belief $\hat{\mu}_m\in\Delta(T')$ and $\hat{\rho}_m\in\mathrm{BR}_m(\hat{\mu}_m)$ such that $u_t(m,\hat{\rho}_m)=u_t(\omega)$ and there is *no* type $t'\in T$ such that $u_{t'}(m,\hat{\rho}_m)>u_{t'}(\omega)$. Let (σ,ρ,μ) be a sequential equilibrium with outcome ω (recall that

ω is assumed to be a sequential outcome). Consider the assessment

$$(\sigma', \rho', \mu') := \big(\sigma, (\hat{\rho}_m, \rho_{-m}), (\hat{\mu}_m, \mu_{-m})\big) .$$

It is easy to see that (σ', ρ', μ') is in $\mathrm{SE}_{\omega,m}(T')$ and is such that m is a best response for t, contradicting our initial assumption. $\square$

Proof of Proposition 3.1

Proof. We begin the proof with a useful result.

Lemma A.1. *For all X, all sets $T', T'' \subset T$ with $T' \supset T''$, and all $t \in T$, if $X_t(T')$ holds then $X_t(T'')$ holds.*

Proof. Take some $T', T'' \subset T$ with $T' \supset T''$, and assume that $X_t(T')$ holds. Observe that $D_t(T') = \mathrm{BR}_m(T') \cap D_t(T)$ and $D_t^0(T') = \mathrm{BR}_m(T') \cap D_t^0(T)$. Note that, if $X \neq \mathrm{D1}$, then $X_t(T')$ holds if and only if

$$\mathrm{BR}_m(T') \cap (D_t(T) \cup D_t^0(T)) \subset \mathrm{BR}_m(T') \cap \mathcal{R}_t^X , \tag{A.1}$$

where $\mathcal{R}_t^{\mathrm{IC}} = \emptyset$, $\mathcal{R}_t^{\mathrm{D2}} = \cup_{t' \in T \setminus \{t\}} D_{t'}(T)$, and $\mathcal{R}_t^{\mathrm{NWBR}} = \cup_{t' \in T} D_{t'}(T)$. It is then clear that, since $\mathrm{BR}_m(T') \supset \mathrm{BR}_m(T'')$, equation (A.1) implies

$$\mathrm{BR}_m(T'') \cap (D_t(T) \cup D_t^0(T)) \subset \mathrm{BR}_m(T'') \cap \mathcal{R}_t^X .$$

Hence, $X_t(T'')$ holds. The argument for $X = \mathrm{D1}$ is analogous. Intuitively, even though shrinking T' to T'' typically shrinks each side of the "$\subset$" in the requirements of each of the conditions for a type to be excludable according to each criterion X provided in Section 3.1, the inclusion persists because when a receiver's best response is eliminated on one side, it is eliminated from the other side as well. $\square$

(End of proof of Lemma A.1. Proof of Proposition 3.1 continues.)

Let $(T_n)_{n \in \mathbb{N}}$ and $(T'_n)_{n \in \mathbb{N}}$ be the sequences of non-excluded types under two implementations of IEX. Assume also, for the sake of contradiction, that $T_\infty \neq T'_\infty$. Without loss of generality, assume that $T'_\infty \setminus T_\infty \neq \emptyset$.[20] For each $t \in T'_\infty \setminus T_\infty$, let $n(t)$ satisfy that $t \in T_{n(t)}$ and $t \notin T_{n(t)+1}$, so $X_t(T_{n(t)})$ holds. Let $\hat{t}$ be such that $n(\hat{t})$ is the minimum of $\{n(t) | t \in T'_\infty \setminus T_\infty\}$, so it must be that $T_{n(\hat{t})} \supset T'_\infty$.[21] Then,

[20] Note that both $(T_n)_{n \in \mathbb{N}}$ and $(T'_n)_{n \in \mathbb{N}}$ are constant for $n \geq |T|$, so T_∞ and T'_∞ are well defined.

[21] Indeed, note that by the definition of $n(\hat{t})$, $T_{n(\hat{t})} \supset T_\infty \cup (T'_\infty \setminus T_\infty) \supset T'_\infty$.

since $T_{n(\hat{t})} \supset T'_\infty$ and $X_t(T_{n(\hat{t})})$ holds, Lemma A.1 implies that $X_t(T'_\infty)$ holds, but this is a contradiction because, by the definition of T'_∞, we have that $X_t(T'_\infty)$ does not hold for all $t \in T'_\infty$, and $\hat{t} \in T'_\infty$ by assumption.

Independence of the order in Definition 3.1. We now prove the claim in the paragraph after Definition 3.1 that, if $T^X_{\omega,m} = \emptyset$, and $\hat{t}, \hat{t}' \in T$ are the last excluded types under two implementations of IEX, then $\mathrm{SE}_{\omega,m}(\{\hat{t}\}) = \emptyset$ if and only if $\mathrm{SE}_{\omega,m}(\{\hat{t}'\}) = \emptyset$. Hence, if ω fails IEX under some implementation of IEX, then ω fails IEX under all implementations of IEX.

Assume $T^X_{\omega,m} = \emptyset$. We use $\mathrm{SET}(X) \subset T^{|T|}$ to denote all (ordered) $|T|$-long *sequences of excluded types* excluded through some implementation of IEX. Assume, for the sake of contradiction, that there are two sequences $(t_n)_{n=1}^{|T|}, (t'_n)_{n=1}^{|T|} \in \mathrm{SET}(X)$ satisfying that $\mathrm{SE}_{\omega,m}(\{t_{|T|}\}) = \emptyset$ and $\mathrm{SE}_{\omega,m}(\{t'_{|T|}\}) \neq \emptyset$. For each n, let T_n and T'_n denote $T \setminus \{t_{n'}\}_{n'=1}^{n-1}$ and $T \setminus \{t'_{n'}\}_{n'=1}^{n-1}$, respectively. Let $\bar{n}$ be such that $t_n = t'_n$ for all $n < \bar{n}$, and such that $t_{\bar{n}} \neq t'_{\bar{n}}$ (hence $T_{\bar{n}-1} = T'_{\bar{n}-1}$ but $T_{\bar{n}} \neq T'_{\bar{n}}$), so we have that $X_{t_{\bar{n}}}(T_{\bar{n}-1})$ and $X_{t'_{\bar{n}}}(T_{\bar{n}-1})$ hold. Assume $\bar{n}$ is *maximal*, that is, there is no pair of implementations $(\hat{t}_n)_{n=1}^{|T|}, (\hat{t}'_n)_{n=1}^{|T|} \in \mathrm{SET}(X)$ satisfying that (i) $\mathrm{SE}_{\omega,m}(\{\hat{t}_{|T|}\}) = \emptyset$, (ii) $\mathrm{SE}_{\omega,m}(\{\hat{t}'_{|T|}\}) \neq \emptyset$, and (iii) $\hat{t}_n = \hat{t}'_n$ for all $n < \bar{n}+1$. Let $\bar{n}' > \bar{n}$ be such that $t'_{\bar{n}'} = t_{\bar{n}}$. There are three cases:

1. Assume first $\bar{n} = |T|-1$. For each $\alpha \in [0,1]$, let $\hat{\mu}^\alpha_m := \alpha \circ t_{\bar{n}} + (1-\alpha) \circ t'_{\bar{n}}$. Note that there is *no* sequential equilibrium $(\sigma, \rho, \mu) \in \mathrm{SE}_{\omega,m}(\{t'_{\bar{n}}\})$ with $\mu_m = \hat{\mu}^1_m$, that there is at least one sequential equilibrium $(\sigma', \rho', \mu') \in \mathrm{SE}_{\omega,m}(\{t_{\bar{n}}\})$ with $\mu'_m = \hat{\mu}^0_m$, and that there is *no* sequential equilibrium $(\sigma'', \rho'', \mu'') \in \mathrm{SE}_{\omega,m}(\{t_{\bar{n}}, t'_{\bar{n}}\})$ with $\mu''_m = \hat{\mu}^\alpha_m$ for any $\alpha \in [0,1]$ where m is a best response for some $t \in T$.[22] We now show this leads to a contradiction. Let α^- be the supremum α for which there is a sequential equilibrium $(\sigma''', \rho''', \mu''') \in \mathrm{SE}_{\omega,m}(\{t_{\bar{n}}, t'_{\bar{n}}\})$ with $\mu'''_m = \hat{\mu}^\alpha_m$ for all $\alpha \in [0, \alpha^-)$. It is then easy to see that there must be some $\rho^- \in \mathrm{BR}_m(\hat{\mu}^{\alpha^-}_m)$ such that $u_t(m, \rho^-) \leq u_t(\omega)$ for all $t \in T$ and $u_t(m, \rho^-) = u_t(\omega)$ for at least one $t \in T$.[23]

2. Assume now that $\bar{n} < |T|-1$ and that $\bar{n}' < |T|$. Note that the sequence $(\check{t}_n)_{n=1}^{|T|}$ defined as

$$(\check{t}_n)_{n=1}^{|T|} := (t'_1, \ldots, t'_{\bar{n}'-2}, t'_{\bar{n}'}, t'_{\bar{n}'-1}, \ldots, , t'_{|T|})$$

is such that $(\check{t}_n)_{n=1}^{|T|} \in \mathrm{SET}(X)$. Indeed, we have that (i) $t'_{\bar{n}'} = t_{\bar{n}}$, (ii) $X_{t_{\bar{n}}}(T'_{\bar{n}})$ holds, (iii) $X_{t'_{\bar{n}'}}(T'_{\bar{n}'})$

[22]Indeed, if such an equilibrium would exist, then the indifferent type t could not have been excluded (if $t \notin \{t_{\bar{n}}, t'_{\bar{n}}\}$) or would be such that $X_t(\{t_{\bar{n}}, t'_{\bar{n}}\})$ does not hold (if $t \in \{t_{\bar{n}}, t'_{\bar{n}}\}$).

[23]This follows from the continuity of $u_t(m, \cdot)$ for all t and the closedness, non-emptyness, upper-hemicontinuity, and convexity of the best response correspondence $\mu_m \mapsto \mathrm{BR}_m(\mu_m)$.

holds, and, (iv) by Lemma 3.1, $X_{t'_{\bar{n}}}(T'_{\bar{n}})$ holds. Proceeding iteratively, we have that the sequence

$$(t''_n)_{n=1}^{|T|} := (t'_1, ..., t'_{\bar{n}-1}, t'_{\bar{n}'}, t'_{\bar{n}}, ..., t'_{\bar{n}'-1}, t'_{\bar{n}'+1}, ..., t'_{|T|})$$

is such that $(t''_n)_{n=1}^{|T|} \in \text{SET}(X)$. This contradicts our assumption that $\bar{n}$ is maximal (in the sense described above), since now $\bar{n}+1$ satisfies the same properties for the sequences $(t'_n)_{n=1}^{|T|}$ and $(t''_n)_{n=1}^{|T|}$.

3. Assume finally that $\bar{n} < |T|-1$ and that $\bar{n}' = |T|$. If $\text{SE}_{\omega,m}(\{t'_{|T|-1}\}) \neq \emptyset$, then the same argument as in the second case can be applied. If, instead, $\text{SE}_{\omega,m}(\{t'_{|T|-1}\}) = \emptyset$, then we define

$$(t''_n)_{n=1}^{|T|} := (t'_1, ..., t'_{|T|-2}, t'_{|T|}, t'_{|T|-1}) ,$$

which satisfies that $(t''_n)_{n=1}^{|T|} \in \text{SET}(X)$ and that $\text{SE}_{\omega,m}(\{t''_{|T|}\}) \neq \emptyset$. This again violates the assumption that $\bar{n}$ is maximal. □

Proof of Proposition 3.2

Proof. Assume that $T_n \neq \emptyset$ and $\text{SE}_{\omega,m}(T_n) = \emptyset$ at some step n of an implementation of IEX. Assume also, for the sake of contradiction, that ω passes IEX; hence, it must be that $n < |T|$. Then, since $T_{n'} \subset T_n$ for all $n' > n$, we have that $\text{SE}_{\omega,m}(T_{n'}) = \emptyset$ for all $n' > n$. By Definition 3.1, this implies that ω fails IEX, a contradiction. □

Proof of Proposition 4.2

Proof. Let ω be a sequentially stable outcome and m an off-path message. We will proceed by induction, showing that if $T' \subset T$ is such that $\text{SE}_{\omega,m}(T') \neq \emptyset$ and $t \in T'$ is such that $X_t(T')$ holds, then either $T' = \{t\}$ (in which case ω passes IEX) or $\text{SE}_{\omega,m}(T' \setminus \{t\}) \neq \emptyset$.

Let $T' \subset T$ be such that $\text{SE}_{\omega,m}(T') \neq \emptyset$ and $t \in T'$ be such that $X_t(T')$ holds. Since that $X_t(T')$ holds implies that $\text{NWBR}_t(T')$ holds (recall expression (3.2)), we have that m is never a best response for t in any element of $\text{SE}_{\omega,m}(T')$. If $T' = \{t\}$ then the result holds. Assume then that $T' \setminus \{t\} \neq \emptyset$. Pick an arbitrary $\mu_m \in \Delta(T')$. By Proposition 4.1, there are some $\alpha \in [0,1]$, $\mu'_m \in \Delta(T)$, and $\rho_m \in \text{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu'_m)$ satisfying that $u_t(m,\rho) \leq u_t(\omega)$ for all t, and if $\alpha \neq 1$, then $u_t(m,\rho_m) = u_t(\omega)$ for all t with $\mu'_m(t) > 0$. Since m is never a best response for t in any element of $\text{SE}_{\omega,m}(T')$, it must be that $\mu'_m \in \Delta(T' \setminus \{t\})$, and hence $\mu''_m \in \Delta(T' \setminus \{t\})$. Letting $(\hat{\sigma}, \hat{\rho}, \hat{\mu})$ be an element of $\text{SE}_{\omega,m}(T')$ (which exists since $\mathrm{SE}_{\omega,m}(T') \neq \emptyset$), it is easy to see that

$$(\hat{\sigma},(\rho_m,\hat{\rho}_{-m}),(\mu''_m,\hat{\mu}_{-m})) \in \mathrm{SE}_{\omega,m}(T' \backslash \{t\})$$

This proves that $\mathrm{SE}_{\omega,m}(T' \backslash \{t\})$ is non empty; hence, the proof is concluded. □

Proof of Proposition 4.3

Proof. Let $T_n \neq \emptyset$ be the set of non-excluded types remaining after (not-necessarily-exhaustively) applying IEX. Fix some $\mu_m \in \Delta(T_n)$. Assume, that there are some $\alpha \in [0,1]$, $\mu'_m \in \Delta(T)$, and $\rho_m \in \mathrm{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu'_m)$, satisfying the condition in Proposition 4.1. Assume also, for the sake of contradiction, that $\mu'_m(t) > 0$ for some $t \notin T_n$. This implies that $\rho_m \in D^0_t(T_n)$ and $\rho_m \notin D_{t'}(T_n)$ for all $t' \in T$, so

$$D^0_t(T_n) \not\subset \cup_{t' \in T \backslash \{t\}} D_{t'}(T_n) \; .$$

Nevertheless, we claim that $\mathrm{NWBR}_t(T_n)$ holds, hence we reach a contradiction. To see why $\mathrm{NWBR}_t(T_n)$ holds, let $T_{n'}$ denot ethe step in IEX where t was excluded. By expression 3.2, we have that $\mathrm{NWBR}_t(T_{n'})$ holds, and since $T_n \subset T_{n'}$, Lemma A.1 implies that $\mathrm{NWBR}_t(T_n)$ holds. □

Proof of Proposition 4.4

Proof. Fix a sequential outcome ω and an off-path message m. We separate the proof into two parts. Throughout, we assume that there are $T' \subset T$ and $\check{\rho}_m \in \mathrm{BR}_m(T')$ such that $u_t(m,\check{\rho}_m) = u_t(\omega)$ for all $t \in T'$ and $u_t(m,\check{\rho}_m) \leq u_t(\omega)$ for all $t \notin T'$. We let $\check{\mu}_m \in \Delta(T')$ be such that $\check{\rho}_m \in \mathrm{BR}_m(\check{\mu}_m)$.

Part (i). We first prove that if ω is sequentially stable then it is sequentially stable in the game where m is not available. The result is trivial since the property in Proposition 4.1 holds for all $m' \neq m$, both in the game where m is available and in the game it is not available.

We now prove that if ω is sequentially stable in the game where m is not available, then it is sequentially stable. Since ω is sequentially stable in the game where m is not available, the property in Proposition 4.1 holds for all $m' \neq m$. Also, we note that the condition in Proposition 4.1 is satisfied for m by choosing, for all $\mu_m \in \Delta(T)$, $\alpha := 0$, $\mu'_m := \check{\mu}_m$, and $\check{\rho}_m := \check{\rho}_m$.

Part (ii). We now prove that $T' \subset T^X_{\omega,m}$. Assume, for the sake of contradiction, that $T' \not\subset T^X_{\omega,m}$, so there is some $t \in T'$ such that $t \notin T^X_{\omega,m}$. Fix some IEX procedure and let n be the first step where a type

$\hat{t} \in T'$ was eliminated, that is, $T' \subset T_n$ and $X_{\hat{t}}(T_n)$ holds. By expression (3.2), we have that $\mathrm{NWBR}_{\hat{t}}(T_n)$ holds. This is a contradiction, since $\check{\rho}_m \in \mathrm{BR}_m(T') \subset \mathrm{BR}_m(T_n)$, so $\check{\rho}_m \in D^0_{\hat{t}}(T_n)$, but $\check{\rho}_m \notin D_{t'}(T_n)$ for any $t' \in T$. $\square$

References

Banks, J. S. and J. Sobel (1987): "Equilibrium selection in signaling games," *Econometrica: Journal of the Econometric Society*, 55, 647–661.

Bernheim, B. D. (1984): "Rationalizable strategic behavior," *Econometrica: Journal of the Econometric Society*, 1007–1028.

Carlsson, H. and S. Dasgupta (1997): "Noise-proof equilibria in two-action signaling games," *Journal of Economic Theory*, 77, 432–460.

Chen, Y., N. Kartik, and J. Sobel (2008): "Selecting cheap-talk equilibria," *Econometrica: Journal of the Econometric Society*, 76, 117–136.

Cho, I.-K. (1987): "A refinement of sequential equilibrium," *Econometrica: Journal of the Econometric Society*, 55, 1367–1389.

Cho, I.-K. and D. M. Kreps (1987): "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 102, 179–221.

Cho, I.-K. and J. Sobel (1990): "Strategic stability and uniqueness in signaling games," *Journal of Economic Theory*, 50, 381–413.

Dilmé, F. (2023a): "Robust Information Transmission," *American Economic Review: Insights*, 5, 111–24.

——— (2023b): "Sequentially stable outcomes," Mimeo.

Farrell, J. (1993): "Meaning and credibility in cheap-talk games," *Games and Economic Behavior*, 5, 514–531.

Grossman, S. J. and M. Perry (1986): "Perfect sequential equilibrium," *Journal of Economic Theory*, 39, 97–119.

Kartik, N. (2009): "Strategic communication with lying costs," *The Review of Economic Studies*, 76, 1359–1395.

Kohlberg, E. and J.-F. Mertens (1986): "On the strategic stability of equilibria," *Econometrica: Journal of the Econometric Society*, 54, 1003–1037.

Kreps, D. M. and R. Wilson (1982): "Sequential Equilibria," *Econometrica: Journal of the Econometric Society*, 50, 863–894.

Mailath, G. J., M. Okuno-Fujiwara, and A. Postlewaite (1993): "Belief-based refinements in signalling games," *Journal of Economic Theory*, 60, 241–276.

McLennan, A. (1985): "Justifiable beliefs in sequential equilibrium," *Econometrica: Journal of the Econometric Society*, 53, 889–904.

Nash, J. F. (1950): "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, 36, 48–49.

Pearce, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica: Journal of the Econometric Society*, 1029–1050.

Selten, R. (1965): "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit: Teil I: Bestimmung des dynamischen Preisgleichgewichts," *Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics*, 121, 301–324.

——— (1975): "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4, 25–55.

Spence, M. (1973): "Job market signaling," *The Quarterly Journal of Economics*, 87, 355–374.