Discussion Paper Series – CRC TR 224

Discussion Paper No. 463
Project B 02

Sequentially Stable Outcomes

Francesc Dilmé [1]

September 2023

[1] University of Bonn, Email: fdilme@uni-bonn.de

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Sequentially Stable Outcomes

Francesc Dilmé*

September 2023

Abstract

This paper introduces and analyzes *sequentially stable outcomes* in extensive games. An outcome ω is sequentially stable if for any $\varepsilon > 0$, any version of the game where players make mistakes with small enough probability has a perfect ε-equilibrium with outcome close to ω. Unlike stable outcomes (Kohlberg and Mertens, 1986), sequentially stable outcomes exist for all finite games and are sequentially rational. If there is a unique sequentially stable outcome, such an outcome is the unique stable outcome of the game's agent normal form. Also, sequentially stable outcomes satisfy versions of forward induction, iterated strict equilibrium dominance, and invariance to simultaneous moves. In signaling games, sequentially stable outcomes pass the standard selection criteria, and when payoffs are generic, they coincide with stable outcomes.

Key words: Sequential stability, stable outcome, signaling games.

JEL classification codes: C72, C73.

---

*University of Bonn. fdilme@uni-bonn.de. I thank Sarah Auster, Pierpaolo Battigalli, Simone Cerreia Vioglio, Srihari Govindan, Johannes Hörner, Stephan Lauermannm Massimo Marinacci, Georg Nöldeke, Klaus Ritzberger, and the seminar participants at the Bocconi University, CERGE-EI in Prague, Hebrew University in Jerusalem, Humboldt University in Berlin, Royal Holloway University of London, Tel Aviv University, University of Bonn, University of Naples, EEA-ESEM 2022 in Milan, and Economic Theory Conference in Honor of George J. Mailath's 65th Birthday, for their useful comments, and Marta Garriga-Massoni for her encouragement and support. This work was funded by a grant from the European Research Council (ERC 949465). Support by the German Research Foundation (DFG) through CRC TR 224 (Project B02) and under Germany's Excellence Strategy – EXC 2126/1– 390838866 and EXC 2047 – 390685813, is gratefully acknowledged.

1 Introduction

Refinements are at the core of the analysis of extensive games. Among them, the concept of a *stable set* (of equilibria) by Kohlberg and Mertens (1986) stands out because of its power (it is stronger than most other common refinements), universality (it exists for all games), properties (it satisfies forward induction, iterated dominance, and invariance), and robustness (it provides predictions independent of small perturbations of the game). Roughly speaking, a set of Nash equilibria is stable if it is minimal with respect to the property that, for any vanishing sequence of normal-form trembles, there is a sequence of equilibria approaching the set.

Despite its theoretical appeal, strategic stability is rarely used as a selection criterion in practice. The main reason is that characterizing and using stable sets of equilibria is difficult. The fact that different equilibria in a stable set may predict different behavior makes it difficult to use them in applications. Additionally, in many extensive games, the number of sets of Nash equilibria that may a priori be stable is large. While forward induction and iterated dominance allow some of them to be ruled out, finding a stable set by ruling out all alternatives is typically infeasible. Proving that a given set of equilibria is stable requires showing that it is minimal with respect to the property that any perturbed version of the game has an equilibrium close by; this is often impractical as well.

An alternative is to look for an analogous single-valued concept, called a *stable outcome*, which is an outcome robust to small normal-form trembles. Kohlberg and Mertens (1986) show that games with generic payoffs have stable outcomes.[1] Nevertheless, this alternative has important limitations. First, many dynamic games of interest have non-generic payoffs (e.g., due to quasilinear preferences, payoff-irrelevant signals, assumed functional forms, or constant discount factors), so stable outcomes cannot be used as a universal equilibrium concept. Second, because there is no guarantee of existence, one cannot prove that a given outcome in a particular game is stable by eliminating alternative candidates. Third, it is difficult to prove directly that a given outcome is stable, as this would require proving that any sequence of trembles has a corresponding sequence of equilibrium outcomes converging to the desired outcome.

We propose an outcome-valued equilibrium concept in the spirit of stable outcomes that has most of the same theoretical appeal, exists for all games, and is easier to use in practice. The key difference is that we look for almost-optimal behavior instead of exactly optimal behavior, in perturbed versions of the game. More formally, an outcome ω is *sequentially stable* if, for any

[1] Okada (1981) had previously defined *strictly perfect equilibria* as those that are stable against all perturbations. However, such equilibria do not exist in many games of interest, even when payoffs are generic.

vanishing sequence of (behavioral) trembles $(\eta_n)_n$ there is a sequence $(\varepsilon_n)_n \to 0$ and a sequence of strategy profiles $(\sigma_n)_n$ with outcomes converging to ω, where each σ_n is a perfect ε_n-equilibrium of the game perturbed according to η_n.[2]

We begin by providing a characterization of *sequential equilibria* (Kreps and Wilson, 1982) as limits of sequentially almost-optimal behavior along vanishing sequences of trembles. We show that a strategy profile σ is part of a sequential equilibrium if and only if there exist a vanishing sequence of trembles (i.e., a vanishing sequence of positive probabilities of making mistakes in the choice of each action), a sequence $(\varepsilon_n)_n \to 0$, and a sequence of corresponding perfect ε_n-equilibria converging to σ. That is, our characterization of sequential equilibria coincides with the definition of *(trembling) perfect equilibrium* (Selten, 1975), except that it relaxes the condition $\varepsilon_n = 0$ to $\varepsilon_n \to 0$. Our definition of sequential stability weakens the notion of stability in an analogous way. An implication is that, since sequentially stable outcomes are robust with respect to *all* vanishing sequences of trembles, sequentially stable outcomes are sequentially rational; that is, they are outcomes of sequential equilibria.

We next establish that all extensive games have at least one sequentially stable outcome. To prove this, we argue that any game has at least one outcome equal to the limit of a sequence of stable outcomes of close-by agent normal-form games with generic payoffs.[3] We then show that such an outcome is in fact sequentially stable. Existence is important not only to ensure that sequential stability is universally applicable, but also to make it possible to find sequentially stable outcomes by eliminating alternatives.

We proceed by showing that sequentially stable outcomes have several desirable properties similar to those of stable outcomes. They satisfy a version of the *never a weak best response* (NWBR) condition (Kohlberg and Mertens, 1986): If ω is sequentially stable and an action a is not a best response in any sequential equilibrium with outcome ω, then ω is sequentially stable in the game where a is eliminated. NWBR is shown to imply versions of both forward induction and iterated

[2]It is easy to see that this definition is equivalent to that in the abstract (see Proposition 3.1), and that sequentially stable outcomes are limits of equilibrium outcomes of games with perturbed payoffs (see Proposition 4.2). Recall that Radner (1980) says a strategy profile is a *perfect ε_n-equilibrium* if, whenever a player assigns to an action a probability higher than that of the tremble, the action is sequentially ε_n-optimal (we extend his definition to games without perfect information, see Definition 2.3).

[3]Because we use behavioral trembles instead of normal-form trembles, each stable outcome of the game's agent normal form is also sequentially stable (Proposition 3.3). We also prove a sort of converse result: If there is a unique sequentially stable outcome, then it is also the unique stable outcome of the game's agent normal form (Proposition 4.5).

strict equilibrium dominance. We also show that the restriction of a sequentially stable outcome to an on-path subgame is sequentially stable in that subgame, and that a subgame with a unique sequential equilibrium outcome can be replaced by that outcome without affecting the set of sequentially stable outcomes. Finally, we prove a weak version of invariance: A sequentially stable outcome remains sequentially stable if simultaneous moves are interchanged. We illustrate how these properties can be used to simplify proving or ruling out the sequential stability of a given outcome.

Lastly, we apply our analysis to signaling games. We show that sequentially stable outcomes pass the Intuitive Criterion, D1, and D2 (Cho and Kreps, 1987). Sequential stability thus has the potential to provide a unified approach to selecting equilibria in signaling games. Additionally, we obtain that a signaling game has a unique sequentially stable outcome if and only if it has a unique stable outcome, and also that the set of sequentially stable outcomes coincides with the set of stable outcomes in signaling games with generic payoffs.

Contribution to the literature: Since the definition of Nash equilibrium (Nash, 1951), many equilibrium concepts have been developed, with the aim of selecting equilibria without undesirable properties. Important examples include subgame-perfect equilibria (Selten, 1965), perfect equilibria (Selten, 1975), proper equilibria (Myerson, 1978), sequential equilibria (Kreps and Wilson, 1982), stable sets (Kohlberg and Mertens, 1986), and perfect Bayesian equilibria (Fudenberg and Tirole, 1991b).[4,5] In signaling games, different selection criteria are used (such as the Intuitive Criterion, D1, and D2, by Cho and Kreps, 1987, or divinity, by Banks and Sobel, 1987). This variety of equilibrium concepts has made it increasingly difficult to compare the predictions of the selected equilibria in different games.

We contribute to the literature by providing a new equilibrium concept suited for use across many applications, making it easier to compare predictions. There are three main reasons for this. First, sequentially stable outcomes constitute a single-valued equilibrium concept, always exist, and pass many plausibility tests (for example, they satisfy sequential rationality and forward induction). Second, the concept of a sequentially stable outcome is stronger than numerous previous equilib-

[4]Sometimes, when equilibrium concepts are not powerful enough as a selection criterion, additional ad-hoc requirements are used, such as the "no signaling what you do not know" and "never dissuaded once convinced" conditions for perfect Bayesian equilibria (see Osborne and Rubinstein, 1994). Alternative restrictions on belief updating off the path of play have been used in Cramton (1985), Rubinstein (1985), Bagwell (1990), and Harrington (1993).

[5]Our analysis does not consider payoff uncertainty, studied in Fudenberg et al. (1988). Recently, Takahashi and Tercieux (2020) have shown the existence of outcomes robust to payoff uncertainty for generic payoffs.

rium concepts, such as subgame-perfect, sequential, or perfect Bayesian equilibrium, and it passes standard selection criteria in signaling games. Sequential stability can therefore be used directly with previous work involving unique solutions: for example, if a game has a unique sequential equilibrium, or a unique equilibrium passing D1, then such a game has a unique sequentially stable outcome. We also show that if a game has a perfect equilibrium outcome, or a stable outcome of its agent normal form, such an outcome is sequentially stable. Third, the properties of sequentially stable outcomes—such as NWBR or forward induction, which are defined and used through natural conditions on the optimality of actions in each information set instead of on the global optimality of complete contingent plans—make it easier to compute them. We provide some examples throughout the paper. In a companion paper, Dilmé (2023), we introduce the (lexicographic) ℓ-numbers as a tool for obtaining and using sequentially stable outcomes without needing to work with vanishing trembles.

The rest of the paper is organized as follows. Section 2 provides the notation for extensive-form games used in the rest of the paper and also defines vanishing trembles and perfect ε-equilibria. Section 3 defines sequentially stable outcomes, relates them to sequential equilibria and stable sets of equilibria, and proves that all games have a sequentially stable outcome. Section 4 obtains properties of sequentially stable outcomes and provides techniques to find them. Section 5 characterizes sequential stability in signaling games and shows that sequentially stable outcomes pass common selection criteria. Finally, Section 6 concludes. The appendix contains the proofs of the results.

2 Basic definitions

2.1 Extensive-form games

We now provide the definition and corresponding notation for an extensive-form game.

An (finite) *extensive-form game* $G := \langle A, H, \mathcal{I}, N, \pi, u \rangle$ has the following components. A finite set of *actions* A. A finite set of *histories* H containing finite sequences of actions satisfying that, for all $(a_j)_{j=1}^J \in H$ with $J>0$, we have $(a_j)_{j=1}^{J-1} \in H$ (hence $\emptyset =: (a_j)_{j=1}^0 \in H$); the set of *terminal histories* is denoted T. An *information partition* $\mathcal{I}$ of the set of non-terminal histories satisfying that there is a partition $\{A^I | I \in \mathcal{I}\}$ of A with the property that, for each $I \in \mathcal{I}$ and $h, h' \in H$, we have (i) $(h, a) \in H$ for some $a \in A^I$ if and only if $h \in I$, and (ii) if $h \in I$ and $h' > h$ then $h' \notin I$.[6] A finite set of *players* $N \not\ni 0$. A

[6]Note that we assume, without loss of generality, that each action only belongs to a unique information set (otherwise, one can rename actions). We use $h' > h$ to indicate that h' succedes h.

player assignment $\iota:\mathcal{I}\to N\cup\{0\}$ assigning each information set to a player or to nature. A *strategy by nature* $\pi:\cup_{I\in\iota^{-1}(\{0\})}A^I\to(0,1]$ satisfying $\sum_{a\in A^I}\pi(a)=1$ for each $I\in\iota^{-1}(\{0\})$. For each player $i\in N$, a *(von Neumann–Morgenstern) payoff function* $u_i:T\to\mathbb{R}$. For convenience, we set $u_0(t)=0$ for all $t\in T$.

A *strategy profile* is a map $\sigma:A\to[0,1]$ such that $\sum_{a\in A^I}\sigma(a)=1$ for all $I\in\mathcal{I}$ (i.e., it is a probability distribution for each set of actions available at each information set) and $\sigma(a)=\pi(a)$ for all a played by nature (i.e., nature plays according to π). We let Σ be the set of strategy profiles.

2.2 Trembles and vanishing trembles

We proceed by defining a tremble and a vanishing tremble. The analysis of trembles and the corresponding perturbed games was initiated by Selten (1975). For each $a\in A$, we let $I^a\in\mathcal{I}$ be the unique information set where a is available (i.e., satisfying $a\in A^{I^a}$).

Definition 2.1. A *(behavioral) tremble* of G is a function $\eta:A\to(0,1]$ satisfying $\sum_{a\in A^I}\eta(a)\leq 1$ for all $I\in\mathcal{I}$ and $\eta(a)\leq\pi(a)$ for all $a\in A$ such that $\iota(I^a)=0$.

As is common, we will interpret $\eta(a)\in(0,1]$ as the smallest probability with which player $\iota(I^a)$ can decide to select action $a\in A$. A tremble thus represents the probability with which players make mistakes. We use $\Sigma(\eta)$ to denote the set of strategy profiles $\sigma\in\Sigma$ satisfying that, for all $a\in A$, $\sigma(a)\geq\eta(a)$. For each tremble, $G(\eta)$ denotes the *perturbed game* defined by G together with the set of strategy profiles $\Sigma(\eta)$. As we will be interested in studying small trembles, we will often work with vanishing trembles.

Definition 2.2. A *vanishing tremble* is a sequence of trembles $(\eta_n)_n$ where $\eta_n(a)\to 0$ for all $a\in A$.

A vanishing tremble $(\eta_n)_n$ generates a sequence of perturbed games $(G(\eta_n))_n$. Such a sequence approaches G (with the set of strategy profiles Σ), in the sense that the sets of strategy profiles $\Sigma(\eta_n)$ approach Σ (under the Hausdorff distance) as $n\to\infty$.[7]

2.3 Perfect ε-equilibria

We now define almost-optimal behavior in a perturbed game. Because a player chooses an action a only if the corresponding information set I^a (i.e., the information set where a is available) is

[7]Our results hold under any normed distances on the spaces of strategy profiles, outcomes, and payoff functions (since all normed distances generate the same topology in $\mathbb{R}^n$). For concreteness, we take the sup-norm.

reached, we will require ε-optimality at each information set given the continuation payoffs.

Fix a tremble η. Note that all information sets are reached with positive probability under any strategy profile $\sigma \in \Sigma(\eta)$. Then, for each action $a \in A$, the expected continuation payoff of the player playing such an action (which is $\iota(I^a) \in N \cup \{0\}$), conditional on the information set where a is played (i.e., I^a) being reached, is uniquely defined. This payoff is

$$u(a|\sigma) := \frac{\sum_{t \in T^a} \Pr^\sigma(t)\, u_{\iota(I^a)}(t)}{\Pr^\sigma(I^a)\, \sigma(a)}\ , \tag{2.1}$$

where $T^a \subset T$ is the set of terminal histories containing a, and where $\Pr^\sigma(\cdot)$ indicates probability under σ. We omit the subindex $\iota(I^a)$ in $u(a|\sigma)$ as it is uniquely determined by a.

The following generalizes the definition in Radner (1980) of perfect ε-equilibrium for games of perfect information to a perturbed game of imperfect information.

Definition 2.3. Fix $\varepsilon > 0$ and η. We say that $\sigma \in \Sigma(\eta)$ is a *perfect ε-equilibrium of $G(\eta)$* if, for all $a \in A$, we have $\sigma(a) > \eta(a)$ only if $u(a|\sigma) \geq u(a'|\sigma) - \varepsilon$ for all $a' \in A^{I^a}$.

The set of perfect ε-equilibria of $G(\eta)$ is denoted by $\Sigma^*_\varepsilon(\eta)$.[8] In a perfect ε-equilibrium, each player behaves ε-optimally in each of her information sets; that is, a player chooses an action a with a probability higher than the trembling probability only if such an action is ε-optimal *conditional* on the corresponding information set being reached. Because $\sigma(a) \geq \eta(a) > 0$ for all $a \in A$ and $\sigma \in \Sigma(\eta)$, the ε-optimality of the strategy can be evaluated in all information sets without the need to specify a belief system, as they all are on path under σ. We will refer to $\Sigma^*_0(\eta)$ as the set of Nash equilibria of $G(\eta)$.

3 Sequentially stable outcomes

In this section, we introduce the concept of a sequentially stable outcome and prove its existence in any game. We also discuss its relationship to the concepts of sequential equilibrium (Kreps and Wilson, 1982) and stable outcome (Kohlberg and Mertens, 1986, see van Damme, 1991, for a textbook treatment).

[8] It may not seem natural to require all players to behave ε-optimally using the same value ε, as such a requirement seems to presume that payoffs are comparable across players. It will nevertheless become clear that our results would follow under a more general definition where ε is allowed to be different for each action.

3.1 Definition of sequentially stable outcomes

Recall that an *outcome* ω (of G) is a probability distribution over terminal histories. We use $\Omega := \Delta(T)$ to denote the set of outcomes. Each strategy profile $\sigma \in \Sigma$ *generates* a unique outcome ω^{σ}, where each $(a_j)_{j=1}^{J} \in T$ is assigned probability $\omega^{\sigma}((a_j)_{j=1}^{J}) := \prod_{j=1}^{J} \sigma(a_j) \in [0,1]$. We now define sequentially stable outcomes, the main object of study of the current paper.

Definition 3.1. An outcome $\omega \in \Omega$ is *sequentially stable* if for any vanishing tremble $(\eta_n)_n$ there are two sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ such that $(\omega^{\sigma_n})_n$ converges to ω.

In words, an outcome is sequentially stable if, for any vanishing tremble, it can be approximated by sequences of perfect epsilon-equilibria of the corresponding perturbed games for some sequence of vanishing epsilons. To give further intuition for what sequential stability requires, we now characterize sequentially stable outcomes without using vanishing trembles.

Proposition 3.1. *An outcome ω is sequentially stable if and only if for all $\varepsilon, \varepsilon' > 0$ there is some $\delta > 0$ such that, if $\|\eta\| < \delta$, then $G(\eta)$ has a perfect ε-equilibrium with outcome ε'-close to ω.*

An outcome ω is then sequentially stable if, for all $\varepsilon > 0$, any slightly perturbed version of G has a perfect ε-equilibrium with outcome close to ω. In other words, a sequentially stable outcome is such that, for any degree of optimality and precision, any perturbed game with small enough tremble has a close-by almost-optimal outcome.

Example 3.1. Figure 1 shows Cho and Kreps (1987)'s *beer–quiche game*, which is used in Kohlberg and Mertens (1986) to illustrate the power of stability. In it, nature first chooses whether player 1's type is strong or weak, then player 1 chooses beer or quiche, and then player 2, observing player 1's choice but not her type, chooses to fight or not. We now prove that the *quiche outcome* ω_{q}—where player 1 chooses $\mathrm{q_s}$ and $\mathrm{q_w}$ and afterward player 2 chooses $\mathrm{n_q}$—is *not* sequentially stable by using an explicit vanishing tremble. Consider a vanishing tremble $(\eta_n)_n$ where $\eta_n(\mathrm{b_s}) = n^{-1}$ and $\eta_n(\mathrm{b_w}) = n^{-2}$, and assume for a contradiction that there are two sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ with $\omega^{\sigma_n} \to \omega_{\mathrm{q}}$. Note that if $\sigma_n(\mathrm{b_w}) = \eta_n(\mathrm{b_w})$ for infinitely many values n, $\sigma_n(\mathrm{f_b}) = \eta_n(\mathrm{f_b})$ for infinitely many values of n as well (since beer becomes an increasingly stronger signal of strong player 1), but this contradicts that $\mathrm{q_s}$ is sequentially ε_n-optimal for large n (as $\sigma_n(\mathrm{q_s})$ must tend to 1 for ω^{σ_n} to approach ω_{q}). Hence, if n is large enough, $\sigma_n(\mathrm{b_w}) > \eta_n(\mathrm{b_w})$, and so $\mathrm{b_w}$ must be sequentially ε_n-optimal. However, since $\sigma_n(\mathrm{n_q}) \to 1$ as $n \to \infty$, the weak player 1's payoff from choosing $\mathrm{q_w}$ (which converges to 3) is larger than the payoff from choosing $\mathrm{b_w}$ (which is at most 2), contradicting that $\mathrm{b_w}$ is sequentially ε_n-optimal for n large enough.

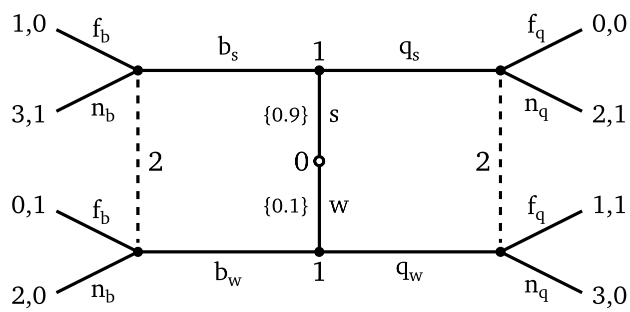

Figure 1

Example 3.2 (continuation of Example 3.1). We now prove that the *beer outcome* ω_b—where player 1 chooses b_s and b_w and afterward player 2 chooses n_b— is sequentially stable. Fix an arbitrary vanishing tremble $(\eta_n)_n$. For $n \in \mathbb{N}$ with $\eta_n(q_s) \geq 9\,\eta_n(q_w)$ define $\sigma_n(a) := \eta_n(a)$ for all $a \in \{q_s, q_w, n_q, f_b\}$, which pins down the value of $\sigma_n(a)$ for all a. Note that, under such σ_n, if n is large enough, each type of player 1 strictly loses from choosing quiche, and that player 2 assigns a probability larger or equal to 0.5 to (s, q_s) at her right information set. For $n \in \mathbb{N}$ with $\eta_n(q_s) < 9\,\eta_n(q_s)$, define $\sigma_n(q_s) := 9\,\eta_n(q_w)$, $\sigma_n(n_q) := 0.5 - \eta_n(f_b)$, and $\sigma_n(a) := \eta_n(a)$ for $a \in \{q_w, f_b\}$, which again pins down the value of $\sigma_n(a)$ for all a. Now, under σ_n and for n is large enough, the strong type strictly loses from choosing quiche, while the weak type is indifferent between beer and quiche. Also, player 2 assigns a probability equal to 0.5 to (s, q_s) at her right information set. It is easy to see that $\sigma_n \in \Sigma_0^*(\eta_n)$ for all n large enough and that $\omega^{\sigma_n} \to \omega_b$; hence, ω_b is sequentially stable.

3.2 Relationship to sequential and stable outcomes

Relationship to sequential outcomes

Part of the motivation for our analysis is to provide an equilibrium concept that is similarly powerful and significantly easier to use than stability. To some degree, this parallels the motivation in Kreps and Wilson (1982) for introducing the concept of *sequential equilibrium*, designed to be similarly powerful but simpler to use than the *(trembling-hand) perfect equilibrium* of Selten (1975): as Kreps and Wilson state, "It is vastly easier to verify that a given equilibrium is sequential than that it is perfect" (p. 264).

Notably, even though the approach of Kreps and Wilson is quite different from ours, their solution and ours end up being based on the same principle: to look for limits of ε-optimal behavior along vanishing trembles instead of limits of exactly optimal behavior. To see this, we now provide a new characterization of sequential outcomes (i.e., outcomes of sequential equilibria) in terms of vanishing trembles and perfect ε-equilibria.

Proposition 3.2. *An outcome ω is sequential if and only if there exist a vanishing tremble $(\eta_n)_n \to 0$, a sequence $(\varepsilon_n)_n \to 0$, and a sequence $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ such that $(\omega^{\sigma_n})_n$ converges to ω.*

Our characterization makes it clear that the concept of sequential equilibrium is a weakening of perfect equilibrium; it requires only ε_n-optimality along the sequence, instead of exact optimality (i.e., $\varepsilon_n = 0$). Analogously, we relax stability by just requiring $\varepsilon_n \to 0$ along the sequence instead of requiring $\varepsilon_n = 0$ for all n (see below for a detailed discussion). As we shall see, relaxing the requirement of exact optimality will provide numerous advantages, as it will ensure the existence of sequentially stable outcomes for all games while preserving many desirable properties. See Blume and Zame (1994) for an analogous characterization of sequential equilibria as limits of Nash equilibria of games with perturbed actions and payoffs, and Myerson and Reny (2020) for an extension and characterization of sequential equilibria to games with infinite sets of signals and actions in terms of perfect conditional ε-equilibria.

We finish this section by noting that the characterization of sequential equilibria in Proposition 3.2 is equivalent to requiring the condition in Definition 3.1 to hold for some vanishing tremble instead of for all of them. As the following corollary states, this implies that sequential stability is a refinement of sequential equilibrium. In other words, a sequentially stable outcome conforms to the requirement of "backward induction" in Kohlberg and Mertens (1986) (van Damme, 1991, calls such a property "sequential rationality", requiring that "any solution contains a sequential equilibrium").

Corollary 3.1. *A sequentially stable outcome is sequential.*

Example 3.3 (continuation of Example 3.2). Example 3.1 shows that the quiche outcome is *not* sequentially stable by explicitly showing that, for a given vanishing tremble, there are no sequences $(\varepsilon_n)_n$ and $(\sigma_n)_n$ satisfying the conditions in Definition 3.1. Such an outcome is nevertheless sequential: It is easy to see that, for any vanishing tremble $(\eta_n)_n$ where $\eta_n(\mathrm{b_s}) = n^{-2}$ and $\eta_n(\mathrm{b_w}) = n^{-1}$, the sequence $(\sigma_n)_n$ pinned down by $\sigma_n(a) := \eta_n(a)$ for all $a \in \{\mathrm{b_s}, \mathrm{b_w}, \mathrm{n_b}, \mathrm{f_q}\}$ satisfies the condition in Definition 3.1 for some $(\varepsilon_n)_n \to 0$; hence, by Proposition 3.2, the quiche outcome is sequential.

Relationship to stable sets and stable outcomes

The definition of sequentially stable outcomes differs from the definitions of stable sets and outcomes in Kohlberg and Mertens (1986) in two important ways. The first difference is that we require ε_n-optimality (for some $\varepsilon_n \to 0$) instead of exact optimality (i.e., $\varepsilon_n = 0$) along the sequence. The second difference is that Kohlberg and Mertens perturb the set of normal-form strategies (i.e., a player's tremble is a full-support distribution over her full contingent plans), while we consider independent trembles to the actions in each information set. We see these two departures as necessary for the goal of obtaining a single-valued equilibrium concept that exists in all games, has high selection power, and possesses desirable properties that permit one to make arguments about the incentive to take actions instead of the incentive to choose full contingent plans (e.g., use the one-shot deviation principle). Let us elaborate.

As we shall see, requiring sequential almost-optimality along the sequence (instead of exact optimality) is a minimal relaxation that maintains important properties while improving tractability and ensuring the existence of sequentially stable outcomes in all games.[9] While such relaxation preserves significant selection power in extensive games, the same approach does not work when applied to normal-form games. To see why, note that Jackson et al. (2012) show that the set of Nash equilibria of a normal-form game coincides with the set of strategy profiles σ satisfying that, for any vanishing tremble $(\eta_n)_n$, there exist two sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n)) \to \sigma$. The implication is that all Nash outcomes of a normal form game are sequentially stable (and hence they are also sequential outcomes).[10]

Given our requirement of approximate sequential rationality along the sequence, we perceive the focus on behavioral vanishing trembles as natural. As Corollary 3.1 establishes, the requirement of stability with respect to behavioral trembles implies that sequential stability satisfies sequential rationality, and we will see that it enables the use of sequential equilibria to characterize sequentially stable outcomes. Similarly, Proposition 3.1 ensures that sequentially stable behavior coincides with behavior that is nearly sequentially optimal for any small enough tremble, which

[9]For example, it permits one to define or verify properties directly "in the limit", just as sequential optimality is only evaluated using the assessment in a sequential equilibrium (Dilmé, 2023, characterizes assessment consistency and sequential stability directly at the limit).

[10]Similarly, Fudenberg and Tirole (1991a, Theorems 14.5 and 14.6) show that, in normal-form games, all outcomes of Nash equilibria are robust to payoff perturbations. Nonetheless, Takahashi and Tercieux (2020) show that requiring outcomes to be robust to payoff perturbations has a significant selection power in extensive-form games. However, like stable outcomes, robust outcomes do not exist for all games and are difficult to compute and use.

gives an additional sense of robustness. In contrast, sequential rationality is not guaranteed in stable sets or stable outcomes. For instance, Kohlberg and Mertens (1986) show that stable sets and outcomes do not satisfy sequential rationality by providing an example of a game with a stable outcome that is not the outcome of the unique sequential equilibrium (see their Figure 11). Furthermore, van Damme (1991) exhibits a game with a stable outcome which is not the outcome of the unique subgame-perfect Nash equilibrium (see his Example 10.3.4).[11] By Corollary 3.1, the unique sequentially stable outcome of the games in these examples is the outcome of their unique sequential equilibrium.

To further connect stability and sequential stability, we say that ω is *extensive-form stable* if satisfies Definition 3.1 with the requirement that $\varepsilon_n = 0$ for all n. It is not difficult to see that ω is an extensive-form stable outcome of G if and only if it is a stable outcome of its agent normal form. While stability neither implies nor is implied by sequential stability or extensive-form stability, we have that extensive-form stability is stronger than sequential stability (but does not satisfy existence). The reason is that if a game coincides with its agent normal form, Kohlberg and Mertens's stability coincides with Definition 3.1 with $\varepsilon_n = 0$ instead of $\varepsilon_n \to 0$. Indeed, as noted before, the set of perfect 0-equilibria coincides with the set of Nash equilibria of a perturbed game $G(\eta)$ because all information sets are reached on path. It is then readily verified that if a game has an extensive-form stable outcome, such otucome is sequentially stable as well.

Proposition 3.3. *Every extensive-form stable outcome is sequentially stable.*

3.3 Existence of sequentially stable outcomes

We now prove the existence of a sequentially stable outcome.

Proposition 3.4. *Any game G has a sequentially stable outcome.*

As the introduction explains, the proof of Proposition 3.4 is divided into two steps. We first argue that any game has an outcome that is the limit of stable outcomes along a sequence of the agent extensive-form game of G with perturbed payoffs. This result follows from the existence of stable outcomes for generic payoffs, which implies that for any generic sequence of payoff functions $(u_m : T \to \mathbb{R}^N)_m$ converging to u, we can find a corresponding sequence of stable outcomes $(\omega_m)_m$,

[11] Mertens (1989) provides a definition of stable sets different from that in Kohlberg and Mertens (1986) and shows that it corrects some undesirable properties of the previous definition (in particular, under his definition, a stable set contains a sequential equilibrium). Mertens's definition is nevertheless remarkably involved and difficult to use in practice.

which can be assumed to converge to some $\omega \in \Omega$. The second step of the proof shows that ω is sequentially stable. In this part, we take a vanishing tremble $(\eta_n)_n$ and fix, for each m, a sequence $(\sigma_{m,n})_n$ with outcomes converging to ω_m where $\sigma_{m,n} \in \Sigma_0^*(\eta_n, u_m)$ for all n (which exists because ω_m is stable), where $\Sigma_\varepsilon^*(\eta_n, u_m)$ indicates the set of perfect ε-equilibria of the game perturbed according to η_n with payoff u_m. We use a standard diagonal argument to prove that there exists an increasing sequence $(n_m)_m$ and a sequence $(\varepsilon_m)_m$ such that each $\sigma_{m,n_m} \in \Sigma_{\varepsilon_m}^*(\eta_{n_m}, u)$ for all m and $\omega^{\sigma_{m,n_m}}$ converges to ω. Given that the argument holds for any vanishing tremble, it follows that ω is a sequentially stable in the agent extensive-form of G, which implies it is also sequentially stable in G. We further show that the correspondence that maps payoffs to the set of sequentially stable outcomes is upper hemicontinuous (see Lemma A.2).

The existence of a sequentially stable outcome in all games contrasts with the fact that stable outcomes only exist for generic payoffs. This negative result motivated Kohlberg and Mertens (1986) to favor a set-valued equilibrium concept, which is more difficult to interpret and use. It is thus clear that the converse of Proposition 3.3 is not true in general: while all games have sequentially stable outcomes, some may not be extensive-form stable. Nevertheless, we will see that a sort of converse result will be true when there is a unique sequentially stable outcome (Proposition 4.5), and will show that sequential stability preserves many of the good properties of stable outcomes.

Example 3.4 (continuation of Example 3.3). Since the beer–quiche game only has two outcomes of sequential equilibria (the quiche outcome and the beer outcome, see Examples 3.1 and 3.2), and since Example 3.1 proves that the quiche outcome is *not* sequentially stable, it follows immediately from Proposition 3.4 that the beer outcome is sequentially stable (hence, in light of Proposition 3.4, the proof in Example 3.2 is pedagogical but not necessary).

4 Properties of sequentially stable outcomes

Kohlberg and Mertens (1986) show that stable sets satisfy forward induction, iterated dominance, admissibility, and invariance. In this section, we show that sequentially stable outcomes have similar properties and explain how they relate to those provided by Kohlberg and Mertens. We also provide some examples showing how they are used.

4.1 Forward induction and iterated strict equilibrium dominance

We begin with a property that is useful for proving or ruling out the sequential stability of outcomes, and that it implies forward induction and iterated strict equilibrium dominance.

Proposition 4.1 (Never a weak best response, NWBR)**.** *Let ω be a sequentially stable outcome. Assume $a \in A$ is not sequentially optimal under any sequential equilibrium with outcome ω. Then ω is a sequentially stable outcome of the game where a is eliminated (as are all histories following it).*[12]

The intuition behind NWBR is the following. Let $\hat{G}$ be the game obtained by eliminating an action a that is not sequentially optimal under any sequential equilibrium with outcome ω. Fix a vanishing tremble in $\hat{G}$, and extend it to a vanishing tremble in G assigning to a a sequence of probabilities vanishing at a much faster rate than the probabilities the vanishing tremble assigns to any other action. Take a corresponding sequence of perfect ε_n-equilibria with outcomes converging to ω (which exists because ω is sequentially stable). The proof shows that the restrictions of perfect ε_n-equilibria to $\hat{G}$ generate a sequence of perfect $\hat{\varepsilon}_n$-equilibria, for some sequence $(\hat{\varepsilon}_n)_n \to 0$. Intuitively, since a is never sequentially optimal, each of the perfect ε_n-equilibria assigns to a the same very low probability as it has in the vanishing tremble if n is high enough. As a result, any history containing a has a much lower likelihood than any history not containing a, ensuring the asymptotic sequential rationality of the restriction to $\hat{G}$.

Our definition of NWBR is analogous to that in Kohlberg and Mertens (1986), but applied to simpler objects. As Fudenberg and Tirole (1991a) explain, Kohlberg and Mertens establish that "a *stable set* contains a stable set of any game obtained by deleting any *strategy* that is not a weak best response to any of the opponents' strategy profiles in the set" (p. 445). In contrast, we determine that a *sequentially stable outcome* is a sequentially stable outcome of any game obtained by deleting any *action* that is not a weak best response to any of the opponents' under any sequential equilibria with that outcome. Our definition of NWBR is thus applied to a single-valued object (outcomes instead of equilibrium sets) and requires simpler conditions (on actions instead of full contingent plans).[13] Furthermore, our definition implies the following versions of iterated strict equilibrium

[12] Note that an action is *sequentially optimal* if player $\iota(I^a)$'s continuation payoff at I^a from playing a (computed using the strategy profile and the belief system) is the maximal continuation payoff player $\iota(I^a)$ can obtain by playing some action in A^{I^a}. Note also that ω is an outcome of the game where an action which is off path under ω is eliminated.

[13] Note that when an action is deleted through NWBR, all strategies where $\iota(I^a)$ plays a at I^a are eliminated. In contrast, Kohlberg and Mertens argue that no single-valued concept satisfies NWBR (when applied to strategies instead of actions), which they see as an argument in favor of using set-valued concepts.

dominance and forward induction.

Corollary 4.1. *Let ω be a sequentially stable outcome. Then the following hold:*

1. ***Forward induction:*** *Assume $I \in \mathcal{I}$ is on path under ω and $a \in A^I$ is such that*

$$\max_{\sigma \in \Sigma_0^*(\omega)} u(a|\sigma) < u(I|\omega) , \tag{4.1}$$

where $u(I|\omega)$ is player $\iota(I)$'s payoff under ω conditional on I being reached, and $\Sigma_0^*(\omega)$ is the set of sequential equilibria with outcome ω. Then, if a is eliminated, ω remains sequentially stable.

2. ***Iterated strict equilibrium dominance:*** *If a strictly equilibrium-dominated action (i.e., an action that is not sequentially optimal under any sequential equilibrium) is eliminated, ω remains sequentially stable.*[14]

Forward induction and iterated strict equilibrium dominance are intuitive and often easier to use than NWBR.[15] Forward induction arguments help to rule out candidates for sequentially stable outcomes by proving they are not sequentially stable in the simpler game obtained after deleting an action available on path but not optimal under any continuation play. Iterated strict equilibrium dominance is applied to the game, not to a particular outcome, and hence can be used to simplify the game before assessing the sequential stability of potential candidate outcomes.

Our property of forward induction, which follows directly from NWBR, permits one to eliminate actions that are available on path but are strictly dominated by not deviating (in the sense of (4.1)). Given our construction using behavioral strategies and the common use of forward induction arguments, we see the requirement that the action is on path as natural.[16] The definition of forward induction in Kohlberg and Mertens (1986), by contrast, permits one to eliminate normal-form strategies (i.e., full contingent plans) that "are an inferior response in all the equilibria of the [stable] set" (p. 1029); this is closer to our NWBR condition.

[14] McLennan (1985) calls an action which is not optimal under any sequential equilibrium *useless*.

[15] Note that Corollary 4.1 also holds if the left side of (4.1) is replaced by $\max_{t \in T^a} u_{\iota(I)}(t)$ (that is, when the outcome's payoff at I is higher than the terminal payoff under any terminal history containing a); this condition is more restrictive but may be easier to verify (since it is not necessary to know $\Sigma_0^*(\omega)$). Similarly, one can weaken iterated strict equilibrium dominance to iterated strict dominance as follows: If $I \in \mathcal{I}$ and $a, a' \in A^I$ are such that $\max_{t \in T^a} u_{\iota(I)}(t) < \min_{t \in T^{a'}} u_{\iota(I)}(t)$, then ω remains sequentially stable if a is eliminated (recall that T^a is the set of terminal histories that contain a).

[16] Cho (1987) defines a refinement of sequential equilibrium, called *forward induction equilibrium*, by requiring a condition similar to (4.1), that is, imposing restrictions on the off-path beliefs after actions that are available on path but strictly dominated by the equilibrium actions, under a conveniently defined set of possible continuation plays.

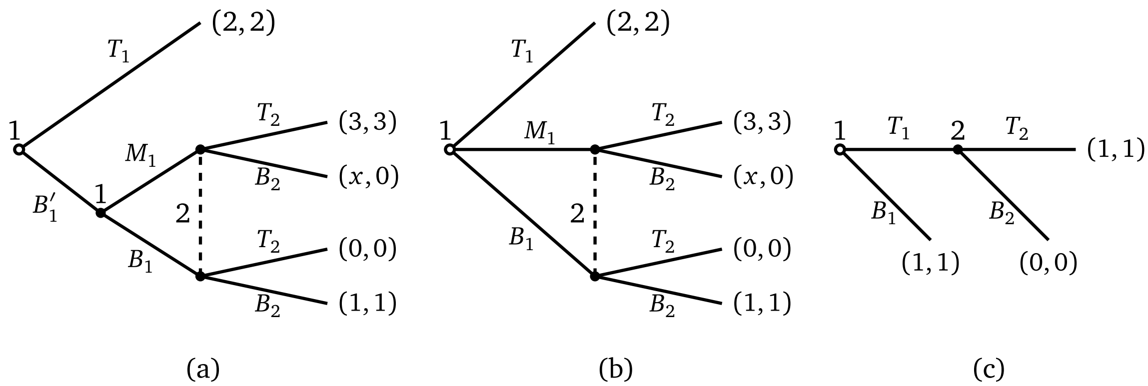

Figure 2

Example 4.1. Before, we argued that stability does not imply sequential stability, as there are games with stable outcomes that are not sequentially rational. Figure 2(a), which coincides with Figure 2 in Kohlberg and Mertens (1986), provides an example showing that sequential stability does not imply stability. In this game, the outcome assigning probability one to T_1 is not stable, but it is sequentially stable when $x < 1$.

Example 4.2 (continuation of Example 3.4). Kohlberg and Mertens argue that the beer outcome ω_{b} is the unique stable outcome of the beer–quiche game (see Figure 1). They do so by claiming that the set of Nash equilibria has two connected components, then ruling out that one is stable using forward induction, and finally claiming that the other connected component must contain a stable set. A complete argument would require proving the existence of two connected components of Nash equilibria and then proving that there are no stable sets with equilibria in both connected sets (note that, according to their definition, a stable set need not be connected). To characterize the possible stable sets, one would then need to further impose minimality.

We now provide a similar but simpler argument, which uses forward induction and equilibrium dominance and is straightforward and complete (see Examples 3.1–3.4 for analogous results using vanishing trembles). Our argument has three steps. (Step 1) We first show that the quiche outcome ω_{q} is *not* sequentially stable. If it were sequentially stable, by forward induction, it would remain sequentially stable upon the elimination of action b_{w}, since the maximum payoff the weak type can achieve by playing it is lower than her payoff under the outcome. In the game without action b_{w}, action f_{b} is strictly (equilibrium) dominated, hence can be eliminated as well. In the resulting game, the strong type prefers playing b_{s} (which can only be followed by n_{b}) to playing q_{s}, so it has no sequential equilibrium with the quiche outcome, contradicting its sequential stability.

(Step 2) We now argue that the beer outcome is the only candidate to be a sequential outcome other than the quiche outcome. To see this, take a sequential equilibrium (σ,μ). It cannot be that $\sigma(\mathrm{b_w})\in(0,1)$, since then $\sigma(\mathrm{n_b})=\sigma(\mathrm{n_q})+1/2$ and the strong type strictly prefers $\mathrm{b_s}$ to $\mathrm{q_s}$, leading to $\sigma(\mathrm{n_b})=1$ and $\sigma(\mathrm{n_q})=0$, so the weak type strictly prefers $\mathrm{b_w}$ to $\mathrm{q_w}$. Similarly, it cannot be that $\sigma(\mathrm{b_s})\in(0,1)$, since then $\sigma(\mathrm{n_b})+1/2=\sigma(\mathrm{n_q})$ and the weak type strictly prefers $\mathrm{q_w}$ to $\mathrm{b_w}$, leading to $\sigma(\mathrm{n_b})=1$ and so $\sigma(\mathrm{n_q})=3/2$. It must then be that $\sigma(\mathrm{b_w}),\sigma(\mathrm{b_s})\in\{0,1\}$. If $\sigma(\mathrm{b_w})\neq\sigma(\mathrm{b_w})$ then the weak type is fought on path while the strong type is not, so the weak type has the incentive to deviate. Hence, the beer outcome is the only candidate to be a sequential outcome. (Step 3) We finally observe that the beer outcome is the unique sequentially stable outcome (by Proposition 3.4), and hence it is the unique stable outcome as well (see Proposition 5.2 below).

4.2 Admissibility and iterated dominance

Kohlberg and Mertens (1986) show that stable sets satisfy admissibility, that is, only contain equilibria where players do not play weakly dominated strategies. It is not difficult to see that a sequentially stable outcome may fail admissibility; that is, players may play weakly a dominated strategy on the path of play. This is not surprising since requiring admissibility and iterated (strict) dominance leads to the non-existence of equilibrium concepts that are not set-valued.[17] We now provide a sense in which the admissibility requirement is fragile to payoff perturbations, but sequentially stable outcomes are not.

See the game in Figure 2(c) for an example of the reasoning behind admissibility. This game has a unique stable outcome, in which player 1 chooses B_1 for sure. This is intuitive since any small tremble by player 2 brings player 1's payoff from choosing T_1 below 1, while choosing B_1 ensures a payoff of 1. However, this argument is fragile to small perturbations on payoffs. Indeed, for any small tremble η, there is a small perturbation in player 1's payoff that makes playing T_1 part of a (unique) Nash equilibrium of the perturbed game. The following proposition establishes a sense in which sequentially stable outcomes are robust: Broadly speaking, a sequentially stable outcome ω is such that any perturbation of the game (in terms of trembles or payoff perturbations) has an equilibrium outcome close to ω.

[17] Kohlberg and Mertens (1986) provide a game in Section 2.7.B (named Ω) that illustrates why requiring iterated dominance leads to the non-existence of a single-valued equilibrium concept. The same example can be used to show that requiring iterated *strict* dominance and additionally requiring admissibility leads to the same non-existence result.

Proposition 4.2. *Let ω be an outcome. The following assertions are equivalent.*

1. *ω is sequentially stable.*
2. *For all $\varepsilon, \varepsilon' > 0$ there are some $\delta, \delta' > 0$ with the property that, for all trembles η with $\|\eta\| < \delta$ and u' with $\|u' - u\| < \delta'$, $G(\eta, u')$ has a perfect ε-equilibrium outcome ε'-close to ω.*
3. *For all $\varepsilon, \varepsilon' > 0$ there is some $\delta > 0$ with the property that, for all trembles η with $\|\eta\| < \delta$, there is some u' with $\|u' - u\| < \varepsilon$ such that $G(\eta, u')$ has a Nash equilibrium outcome ε'-close to ω.*

4.3 Sequential stability in subgames

Selten (1965) introduced the concept of subgame perfect (Nash) equilibrium to give plausibility to equilibrium behavior: Even if the players find themselves off-path, they should continue playing mutual best responses. Sequential rationality has since been a crucial property of some equilibrium concepts (e.g., perfect equilibrium) or a requirement in others (e.g., sequential equilibrium). In addition to adding plausibility, subgame perfection eases the study of games as it permits the use of backward induction. For example, by iteratively replacing subgames with one of their Nash equilibria, one can obtain (subgame perfect) Nash equilibria of the original game. Analyzing each simpler subgame separately is often easier than studying the whole game at once.

The credibility of off-path behavior is a crucial aspect of sequential stability: Since all information sets are on-path for each tremble, requiring stability to all vanishing trembles provides a strong sense of sequential rationality. As a result, as we have shown, sequentially stable outcomes are sequentially rational; that is, they are the outcome of sequential equilibria, which are themselves subgame perfect.[18] Still, because a sequentially stable outcome does not specify the off-path behavior, subgame perfection cannot be applied directly. As Example 3.2 shows, the limiting off-path behavior of the sequence of perfect ε_n-equilibria supporting the outcome may depend on the particular vanishing tremble used. We now provide results that help compartmentalize the game to ease the study of sequential stability.

The following proposition establishes three results. The first says that a subgame with a unique sequential outcome can be replaced by such an outcome without altering the set of sequentially stable outcomes. This result is useful as it permits iteratively reducing the complexity of a game. The second result establishes that if an outcome ω is sequentially stable in the game

[18] In fact, Proposition 3.2 ensures that, if ω is sequentially stable, then for any vanishing tremble $(\eta_n)_n$, any corresponding sequence $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ supporting ω (for some $\varepsilon_n \to 0$) has a subsequence supporting to a sequential equilibrium.

resulting from replacing a subgame with one of its sequentially stable outcomes, then ω is also a sequentially stable outcome of the original game. This result helps find sequentially stable outcomes, as subgames can be iteratively replaced by one of their sequentially stable outcomes. Finally, the third result establishes that the conditional distribution that a sequentially stable outcome induces in an on-path subgame is itself a sequentially stable outcome of the subgame. This helps rule out the sequential stability of a candidate outcome (by arguing that its continuation outcome is not sequentially stable in some on-path subgame) and reduce the possible on-path behavior of sequentially stable outcomes. Overall, these results help find sequentially stable outcomes and prove or disprove the sequential stability of a given outcome.

Proposition 4.3. 1. *Let G' be a subgame of G with a unique sequential outcome ω'. Then the game where G' is replaced by ω' has the same set of sequentially stable outcomes as G.*[19]

2. *Let G' be a subgame of G and ω' a sequentially stable outcome of G'. Let ω be a sequentially stable outcome of the game where G' is replaced by ω'. Then ω is sequentially stable in G.*

3. *Let ω be sequentially stable and let G' be a subgame of G that occurs on the path of ω. Then, the conditional distribution of the terminal histories in G' is a sequentially stable outcome of G'.*

4.4 Invariance

Our focus on behavioral trembles permits us to state definitions (e.g., those of a perfect ε-equilibrium or sequentially stable outcome) and their properties (e.g., NWBR, iterated strict equilibrium dominance) in terms of actions instead of normal-form strategies (i.e., probability distributions over full contingent plans). In extensive games, reasoning in terms of the players' incentives to take actions in each information set is often easier and more natural than reasoning using normal-form strategies, as the latter may be highly complex. An implication of our approach is that, like other equilibrium concepts based on behavioral trembles (e.g., perfect and sequential equilibrium), sequential stability is not invariant to changes in the game tree that preserve the normal form game.

Even though sequential stability does not satisfy invariance, it relies less on the particular game tree of a given normal-form game than other equilibrium concepts. Take, for example, games (a) and (b) in Figure 2 with $x \in (1,2)$, which Kohlberg and Mertens (1986) use to exemplify the invariance of stability (see their Figures 2 and 3). Note that game (b) is obtained by "coalescing"

[19] In the statement, "the game where G' is replaced by ω'" is a game where, at the node where G' is initiated, nature chooses each of the terminal histories t' in the support of ω' with probability $\omega'(t')$. Note that Govindan (1996) proves a result similar to our second result but for stable sets of equilibria.

two moves of player 1 in game (a). Kohlberg and Mertens argue that the outcome ω_1 assigning probability one to T_1 is both a perfect and a sequential equilibrium outcome in game (b) but not in game (a)—even though games (a) and (b) have the same reduced normal form—, while ω_1 is not stable in either game. Their logic can be used to argue that ω_1 is not a sequentially stable outcome of games (a) and (b) (note that B_1 is a strictly dominated action in both games), so the only sequentially stable outcome assigns probability one to (M_1, T_2). Nevertheless, as explained in Example 4.1, sequential stability does not satisfy invariance when $x \in (0,1)$, since in this case ω_1 is sequentially stable (and sequential) in (a) but not in (b).

While there is disagreement on the desirability of invariance as a requirement for equilibrium concepts, most authors would agree that invariance to interchanging simultaneous moves is a basic and necessary requirement (which is also satisfied by stable sets). Indeed, while a modeler of a particular economic activity may identify the sequence of moves to determine the game tree accordingly, she has freedom in how to encode simultaneous moves. The following property states that the set of sequentially stable outcomes does not depend on the particular choice of the sequence of moves.

Proposition 4.4 (Invariance to interchanging simultaneous moves)**.** *Let ω be a sequentially stable outcome. If two information sets I and I' are such that $I' = I \times A^I$ (i.e., I and I' are simultaneous), a game where the order of I and I' is reversed has a sequentially stable outcome equivalent to ω.*

4.5 Approaches to obtaining sequentially stable outcomes

We now discuss procedures that permit one to obtain sequentially stable outcomes in extensive games beyond explicitly showing sequential stability for all vanishing trembles (as in Example 3.2). These procedures partially overcome some of the complications that make using stable sets and stable outcomes difficult in practice, as described in the introduction. After that, we discuss how these results can be used to obtain stable outcomes in some cases.

Through requiring necessary conditions: The first procedure consists in eliminating all candidates for sequentially stable outcomes except for one by using necessary conditions such as the properties established in Propositions 4.1 and 4.3 and Corollary 4.1. Recall that, by Corollary 3.1, only outcomes of sequential equilibria can be sequentially stable. Since sequential equilibria may sometimes be difficult to compute, the procedure of eliminating candidates can be applied to outcomes of a weakening of sequential equilibria, such as perfect Bayesian equilibria (Fudenberg and Tirole, 1991b). Example 4.2 illustrates this technique.

While NWBR and forward induction can be used to rule out the sequential stability of specific outcomes, iterated strict equilibrium dominance permits the direct elimination of "implausible" moves, which simplifies the analysis (note that parts 1 and 2 of Proposition 4.3 also enable one to simplify the game). For instance, the elimination of a strictly dominated action does not change the set of sequentially stable outcomes (see Footnote 15) but typically reduces the set of candidates (e.g., by reducing the set of outcomes of sequential equilibria). It is important to note that if the game resulting from the elimination of a strictly dominated action has a *unique* sequentially stable outcome, then it is the unique sequentially stable outcome of the original game. A similar argument is difficult to make when using iterated dominance as in Kohlberg and Mertens (1986)—which states that "a stable set contains a stable set of any game obtained by deletion of a dominated strategy", see their Proposition 6—as the uniqueness of a stable set in the resulting game only implies that this set is part of a stable set of the original game.

Through a vanishing tremble: The second procedure consists in reducing the field of candidates for sequentially stable outcomes by considering particular vanishing trembles. If one can find a vanishing tremble such that all corresponding sequences of almost-optimal behavior have the same limit outcome ω, then, by the existence of sequentially stable outcomes, ω has to be the unique sequentially stable outcome. The advantage of this approach is that it does not require ruling out the sequential stability of all but one outcome; rather, it lets one prove immediately that a given outcome is the unique sequentially stable outcome. The disadvantage is that it may sometimes be difficult to find the right vanishing tremble and prove that ω is the unique limit outcome.[20,21] Example 5.1 illustrates this technique. See also Examples 3.1 and 5.2 for the use of a vanishing tremble to rule out the sequential stability of a given outcome.

Elimination through a vanishing tremble is particularly convenient in games where the payoff from taking certain actions depends on (and hence communicates) private information, and this payoff satisfies a "single-crossing" condition (e.g., in signaling games, or bargaining games with private information). In these games, even when payoffs are not "generic", there is often a very small set of equilibrium outcomes for perturbed versions of the game where the highest type (i.e., the

[20]We provide a convenient characterization in Lemma A.1 in the Appendix: We show that an outcome $\omega \in \Omega$ is sequentially stable if and only if the property in Definition 3.1 holds for some subsequence of $(\eta_n)_n$ instead of the whole sequence.

[21]In a companion paper (Dilmé, 2023), we define ℓ-numbers as a way to work with limit likelihoods of actions and histories. The advantage of using ℓ-numbers is that the sequential stability of a given outcome can be proved without using sequences of strategy profiles; it is only necessary to verify sequential optimality at the limit.

type that other types want to mimic) trembles more than the low types. Hence, in many cases, vanishing trembles where high types tremble asymptotically more than low types have a unique limit equilibrium outcome, which is then the unique sequentially stable outcome. In signaling games à la Spence (1973), for example, such a stable outcome is often the least costly, fully separating outcome, called the *Riley outcome* (see Riley, 1979), but not necessarily (see Example 5.1).

Combining techniques: Note that the previous procedures for obtaining sequentially stable outcomes (through requiring necessary conditions and through a vanishing tremble) are not mutually exclusive. On the contrary, they can be combined. For example, iterated strict equilibrium dominance can be used to simplify the game. Then, for a given candidate outcome, one can use NWBR, forward induction, or a vanishing tremble to show that the outcome is not sequentially stable in the simplified game, and hence it is not sequentially stable in the original game. If there remains only one sequential outcome, then this is the unique sequentially stable outcome of the original game.

Finding extensive-form stable outcomes

We now provide an important result connecting sequential stability and extensive-form stability (as defined in Section 3.1).

Proposition 4.5. *If there is a unique sequentially stable outcome, it is the unique extensive-form stable outcome.*

Proposition 4.5 establishes that, while sequential stability is weaker than extensive-form stability (by Proposition 3.3), the two concepts coincide when there is a unique sequentially stable outcome. An equivalent result states that if G has a unique sequentially stable outcome, it is the unique stable outcome of its agent normal form. The result is then useful since, as we argued before, it is in general easier to prove that a game has a unique sequentially stable outcome than to prove that it has a unique extensive-form stable outcome. The intuition for the result is as follows. If there is a unique sequentially stable outcome ω, either this outcome is the unique extensive-form stable outcome (in which case the result holds), or the game has no extensive-form stable outcome. Proceeding by contradiction, we assume the second case; hence, there is a vanishing tremble with no corresponding sequence of Nash equilibria converging to ω. Combining this vanishing tremble with a sequence of perturbations of payoffs, we are able to construct a sequence of sequentially stable outcomes of nearby games converging to an outcome ω' different from ω. However, as argued in the proof of Proposition 3.4, ω' must hen be sequentially stable, contradicting that ω is the unique sequentially stable outcome.

5 Sequential stability in signaling games

Since the introduction of signaling games by Spence (1973), many selection criteria have been suggested to address their inherent multiplicity of equilibria. Many such selection criteria are specific to signaling games and difficult to generalize to other classes of games; examples include the Intuitive Criterion, D1, and D2 (Banks and Sobel, 1987) and divinity (Cho and Kreps, 1987). In this section, we relate selection criteria in signaling games to sequential stability.

5.1 Signaling games and sequential stability

A signaling game G^{sig} proceeds as follows. First, nature chooses a type $\theta \in \Theta$ with distribution $\pi \in \Delta(\Theta)$. Having observed θ, the sender chooses a message $m \in M_\theta \subset M$. Finally, having observed the message but not the type, the receiver chooses a response $r \in R_m \subset R$. We assume Θ, M, and R are finite sets. As usual, we let Θ_m be the types who can send message m.[22] Abusing notation, we let $u_\theta(m,r)$ and $u_{\text{r}}(\theta,m,r)$ denote the payoffs of the sender and the receiver, respectively, at $(\theta,m,r) \in T$, and we let $u_\theta(\omega)$ denote the sender's payoff under outcome ω conditional on the realized type being θ. Also, we let $\text{BR}_m(\mu_m) \subset \Delta(R_m)$ be the set of (mixed) best responses of the receiver to message m when the belief about types is $\mu_m \in \Delta(\Theta_m)$, and $\text{BR}_m := \cup_{\mu_m \in \Delta(\Theta_m)} \text{BR}_m(\mu_m)$. Note that G^{sig} is such that players play at most once in each path of play.

The following is a characterization of the set of sequentially stable outcomes of G^{sig}.

Proposition 5.1. *ω is sequentially stable if and only if it is the outcome of a sequential equilibrium and, for any off-path $m \in M$ and $\mu_m \in \Delta(\Theta_m)$, there are some $\alpha \in [0,1]$, $\mu'_m \in \Delta(\Theta_m)$, and $\rho \in \text{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu'_m)$, satisfying that $u_\theta(m,\rho) \le u_\theta(\omega)$ for all $\theta \in \Theta_m$ and, if $\alpha \neq 1$, that $u_\theta(m,\rho) = u_\theta(\omega)$ for all $\theta \in \Theta_m$ with $\mu'_m(\theta) > 0$.*

Banks and Sobel (1987) and Cho and Kreps (1987) find that, in a signaling game with generic payoffs, an outcome satisfies (similar but slightly more complicated than the) conditions in Proposition 5.1 if and only if it is stable (see their Theorem 3 and Proposition 4, respectively). Consequently, the following is an immediate corollary of our Propositions 3.3 and 5.1:

Corollary 5.1. *Generically in payoffs, the set of stable outcomes and the set of sequentially stable outcomes of G^{sig} coincide.*

[22]Note that we abuse notation by using m as a message that can be sent by different types of the sender, given that our definition of an extensive game requires that each action is only played in a unique information set.

We end this section with a result that enables one to show that an outcome of G^{sig} is stable by proving it is its unique sequentially stable outcome, or the other way around.

Proposition 5.2. *An outcome is the unique sequentially stable outcome of* G^{sig} *if and only if it is its unique stable outcome.*

The proof first shows that, in signaling games, if ω is extensive-form stable, then it is stable.[23] It then shows that if ω is stable, then it is sequentially stable. The implication is that if G^{sig} has a unique sequentially stable outcome, then (i) such an outcome is extensive-form stable (by Proposition 4.5) and hence stable, and (ii) there is no other stable outcome, since otherwise, such an outcome would be sequentially stable.

5.2 Signaling refinements

In Cho and Kreps (1987), distinct criteria were proposed to select equilibria in signaling games. These criteria have significantly shaped the handling of equilibrium multiplicity across diverse applications. In this section, we argue that sequential stability is stronger than their refinements.

For conciseness, we refer to D1, D2, and NWBR_{CK} (i.e., Cho and Kreps's version of NWBR; see below) as the *classic selection criteria* (for signaling games).[24] They are based on the following procedure. First, fix an outcome of a sequential equilibrium. Then, for each off-path message, one prunes out all types deemed implausible according to the criterion. Finally, if not all types are pruned, one checks whether there is a sequential equilibrium where, if the sender chooses an off-path message, the receiver assigns probability zero to the pruned-out types. If one such sequential equilibrium exists, the outcome passes the criterion; otherwise, it fails it. The reader may benefit from reading Cho and Kreps's work for intuition and motivation for each selection criterion.

We now show that sequentially stable outcomes pass the classic selection criteria. An implication is that, for each classic selection criterion, there is an outcome passing it. Additionally, if there is a unique outcome passing one of the classic selection criteria, such an outcome is the unique sequentially stable (and stable) outcome of G^{sig}.

[23]The proof of this result is not trivial because the trembles of the reduced-form game (used to determine stability) affect all types of the sender equally and are potentially correlated across types, while the size of the behavioral trembles (used to determine extensive-form stability) may depend on the type and are uncorrelated across types.

[24]We omit the Intuitive Criterion and Banks and Sobel (1987)'s divinity and universal divinity since they are based on a different methodology. Nevertheless, it is not difficult to see (using arguments similar to those in the proof of Proposition 5.3) that if ω is sequentially stable then it passes the Intuitive Criterion.

Proposition 5.3. *Let ω be a sequentially stable outcome of G^{sig} and m an off-path message. Let $\hat{\Theta}\subset\Theta_m$ be the set of all types $\theta\in\Theta_m$ satisfying one of the following conditions:*

(D1) $\exists\theta'\in\Theta_m\backslash\{\theta\}\ \ \forall\rho\in\text{BR}_m\ \ u_\theta(m,\rho)\geq u_\theta(\omega)\Rightarrow u_{\theta'}(m,\rho)>u_{\theta'}(\omega)$*, or*

(D2) $\forall\rho\in\text{BR}_m\ \ \exists\theta'\in\Theta_m\backslash\{\theta\}\ \ u_\theta(m,\rho)\geq u_\theta(\omega)\Rightarrow u_{\theta'}(m,\rho)>u_{\theta'}(\omega)$*, or*

(NWBR$_{CK}$) $\forall\rho\in\text{BR}_m\ \ \exists\theta'\in\Theta_m\backslash\{\theta\}\ \ u_\theta(m,\rho)=u_\theta(\omega)\Rightarrow u_{\theta'}(m,\rho)>u_{\theta'}(\omega)$.

Then, if $\hat{\Theta}\neq\Theta_m$, there is a sequential equilibrium (σ,μ) with outcome ω where $\mu_m(\hat{\Theta})=0$.

5.3 Examples

In this section, we provide two examples. Example 5.1 illustrates how sequential stability helps select outcomes in a standard signaling game where a single-crossing condition applies and shows that the selected outcome may fail to be the Riley outcome. Example 5.2 illustrates how sequential stability can be used in a signaling game not satisfying single crossing, where other selection criteria cannot be used. Note also that since the beer–quiche game studied in Examples 3.1–3.4 and 4.2 is a signaling game, we can use Section 5's results to analyze it further. See Dilmé (2023) for further examples.

Example 5.1 (signaling with single crossing). In this example we consider a version of the Spence (1973) model. Nature first decides the type of player 1, $\theta\in\{\theta_0,\theta_1\}$, with $\pi(\theta_0):=3/4$. Then player 1 chooses the message (effort) $m\in M:=\{0,\Delta,2\Delta,...,\lfloor\Delta^{-1}\rfloor\Delta\}$, for some small $\Delta>0$. Finally, after observing the effort, player 2 decides the response, $r\in\{0,1\}$. The payoffs are[25]

$$u_\theta(m,r):=r-c_\theta\, m \ \text{ and } \ u_{\text{r}}(\theta,m,r):=r\,(2\,\mathbb{I}_{\theta=\theta_1}-1)\,, \tag{5.1}$$

with $1<c_{\theta_1}<c_{\theta_0}<1/\Delta$. Note that the receiver would want to choose $r=0$ when $\theta=\theta_0$ and $r=1$ when $\theta=\theta_1$, while both types would want to choose a low message and want the receiver to choose $r=1$. We let $\overline{m}$ be the smallest message bigger than $1/c_{\theta_0}$. We consider the following vanishing tremble $(\eta_n)_n$: Types θ_0 and θ_1 tremble to all messages with likelihoods n^{-2} and n^{-1}, respectively; that is, $\eta_n(m|\theta_0)=n^{-2}$ and $\eta_n(m|\theta_1)=n^{-1}$ for all m, while the receiver trembles to all responses with probability n^{-1}.

Let $(\varepsilon_n)_n\to 0$ and $(\sigma_n\in\Sigma^*_{\varepsilon_n}(\eta_n))_n$ be such that $(\sigma_n)_n$ supports an assessment (σ,μ) (which by Proposition 3.2 is a sequential equilibrium). Let $m_+<1$ be the highest effort such that $\sigma_n(m|\theta_0)>\eta_n(m|\theta_0)$ for an infinite number of $n\in\mathbb{N}$. It must then be that $\sigma_n(r=1|m)\to 1$ for all $m>m_+$.

[25] For any predicate P, we have $\mathbb{I}_P:=1$ if P is true and $\mathbb{I}_P:=0$ if P is false.

Since, for type θ_0, choosing message 0 strictly dominates choosing any message $m \geq \overline{m}$, we have that $m_+ < \overline{m}$. It must be that the low type prefers not to deviate to choose $m_+ + \Delta$, that is,

$$\sigma_n(r=1|m_+) - c_{\theta_0} m_+ \geq 1 - (m_+ + \Delta) c_{\theta_0} - \varepsilon_n \;\Rightarrow\; c_{\theta_0} \geq \frac{1-\sigma_n(r=1|m_+)+\varepsilon_n}{\Delta} .$$

Since m_+ is optimal for type θ_0, the usual single-crossing property implies that $\sigma_n(m|\theta_1) = \eta_n(m|\theta_1)$ for all $m < m_+$ if n is large enough.[26] Since type θ_0 has to assign positive probability to at least one effort below m_+ in the limit, and since player 2 chooses $r=0$ if the posterior about the type being θ_1 is small enough, type θ_0 must choose $m=0$ with positive probability in the limit. It must then be that

$$m_+ = \lfloor (c_{\theta_0} \Delta)^{-1} \rfloor \Delta \text{ and } \sigma(r=1|m) = c_{\theta_0} m \;\; \forall m \leq m_+ .$$

Generically in c_{θ_1}, there are then two cases.[27] The first case is when $\lfloor (c_{\theta_0} \Delta)^{-1} \rfloor c_{\theta_0} \Delta < 1 - c_{\theta_1} \Delta$, hence type θ_1 strictly prefers $m_+ + \Delta$ to m_+. This implies that σ is uniquely determined by

$$\big(\sigma(m|\theta_0), \sigma(m|\theta_1), \sigma(r=1|m)\big) = \begin{cases} (1,0,0) & \text{if } m=0, \\ (0,0,c_{\theta_0} m) & \text{if } 0<m\leq m_+, \\ (0,1,1) & \text{if } m=m_+ + \Delta, \\ (0,0,1) & \text{if } m > m_+, \end{cases}$$

Hence, there is a unique candidate sequentially stable outcome is the Riley outcome. In this outcome, the low type chooses the least costly message, while the high type chooses the cheapest message that allows full separation. Since the Riley outcome is the unique limit equilibrium outcome for the considered vanishing tremble, it must be both sequentially stable and stable. The second case is when $\lfloor (c_{\theta_0} \Delta)^{-1} \rfloor c_{\theta_0} \Delta > 1 - c_{\theta_1} \Delta$, hence type θ_1 strictly prefers m_+ to $m_+ + \Delta$. This

[26] Here, the single-crossing property says that "if θ_0 (weakly) prefers m_+ to $m < m_+$, then type θ_1 strictly prefers m_+ to m" (this holds because $c_{\theta_1} < c_{\theta_0}$). So, for all $m < m_+$, $\lim_{n\to\infty}(u_{\theta_1}(m_+|\sigma_n) - u_{\theta_1}(m|\sigma_n)) > 0$, hence $\sigma_n(m|\theta_1) = \eta_n(m|\theta_1)$ if n is large enough.

[27] We require that $\lfloor (c_{\theta_0} \Delta)^{-1} \rfloor c_{\theta_0} \Delta \neq 1 - c_{\theta_1} \Delta$, since otherwise there is a spurious multiplicity of limit equilibrium outcomes. Note that, while our specification is standard, it is also highly non-generic, both because of the structure of the message space and the payoffs (5.1). So, stable outcomes cannot be assumed to exist.

implies that σ is uniquely determined by

$$\big(\sigma_n(m|\theta_0),\sigma_n(m|\theta_1),\sigma_n(r=1|m)\big)\to\begin{cases}(2/3,0,0) & \text{if } m=0,\\ (0,0,c_{\theta_0}\,m) & \text{if } 0<m<m_+,\\ (1/3,1,c_{\theta_0}\,m_+) & \text{if } m=m_+,\\ (0,0,1) & \text{if } m>m_+,\end{cases}$$

In this second case, there is again a unique candidate to sequentially stable outcome, but it does not coincide with the Riley outcome. In this outcome, the low type randomizes between the lowest message and a separating message, while the high type chooses the separating message with probability one. Again, this outcome is both sequentially stable and stable.

Example 5.2 (signaling without single crossing). We now present an example of a signaling game where no single-crossing condition holds. Consider two types, $\Theta:=\{\theta_0=0,\theta_1=1\}$, two messages, $M:=\{m_0,m_1\}$, and actions in a grid, $R:=\{0,1/\overline{r},...,1-1/\overline{r},1\}$, for some large even number $\overline{r}/2\in\mathbb{N}$ (so $1/2\in R$). Nature chooses $\theta=\theta_1$ with probability $1/2$. The receiver's payoff is $-(r-\theta)^2$; that is, he "tries to match" the belief about type 1. We also assume that message m_1 is costly, that type 0 prefers high actions, and that type θ_1 prefers intermediate actions:

$$u_{\theta_0}(m,r):=\mathbb{I}_{r\in[1/3,1]}-\mathbb{I}_{m=m_1}\,c_0 \;\text{ and }\; u_{\theta_1}(m,r):=\mathbb{I}_{r\in[1/4,3/4]}-\mathbb{I}_{m=m_1}\,c_1\;,$$

where $c_0,c_1\in(0,1)$. Consider an outcome where both types choose m_1, and where the receiver chooses $r=1/2$. Such an outcome passes all standard selection criteria (D1, D2, and NWBR_{CK}). The reason is that the set of receiver's actions that make deviating profitable for each type are not ordered by inclusion. To prove that the outcome is *not* sequentially stable, consider a tremble where the high type trembles to m_0 with a higher likelihood than the low type, say $\eta_n(m_0|\theta_1):=n^{-1}$ and $\eta_n(m_0|\theta_0):=n^{-2}$. Assume there are two sequences $(\varepsilon_n)_n$ and $(\sigma_n)_n$ with the properties in Definition 3.1. Note that type θ_0's payoff from choosing m_0 has to be asymptotically the same as his payoff from choosing m_1, since otherwise the receiver would assign an increasingly high posterior to type θ_1 after m_1, leading him to choose $r=1$ and making type θ_0 strictly willing to deviate. As a result, the probability with which the receiver has to play an action in $[1/3,1]$ after m_0 should tend to $1/2$ as $n\to\infty$. Note that, for n large enough, the receiver chooses an action r with positive probability after m_0 only if $|\mu_m(\theta_1)-r|<1/\overline{r}$. Hence, letting $k\in\mathbb{N}$ be such that $k/\overline{r}\leq 1/3<(k+1)/\overline{r}$, we have that the receiver puts an increasingly higher probability on $\{k/\overline{r},(k+1)/\overline{r}\}$ after m_0 as n increases. Nevertheless, type θ_1 strictly wants to deviate to m_0, since by doing so she obtains 1 instead of

$1 - c_1$ which is a contradiction. It is not difficult to see that the game has a unique sequentially stable outcome where both types choose m_0 for sure. Hence, such an outcome is stable as well.

6 Conclusions

We have investigated the limits of near-optimal behavior along sequences of perturbed games. When convergence is required along *some* vanishing tremble, sequential outcomes are obtained. When instead convergence is required along *all* vanishing trembles, sequentially stable outcomes are obtained. As sequential equilibria have been extensively studied, our analysis has focused on characterizing sequentially stable outcomes.

We have shown that sequential stability preserves versions of many desirable properties of stability. First, a sequentially stable outcome gives robust predictions: Any perturbation of the game has almost-optimal behavior close to a sequentially stable outcome. Second, a sequentially stable outcome satisfies various plausible requirements: It is the outcome of a sequential equilibrium, and it remains sequentially stable after the elimination of strictly dominated actions or the interchange of simultaneous moves. Finally, sequentially stable outcomes exist in all games and imply most refinements or selection criteria; hence, they can be used to select and compare equilibria across games.

The existence of sequentially stable outcomes for all games facilitates their use in practice. Sequentially stable outcomes can be obtained by ruling out the alternatives through some vanishing tremble, using properties such as never-a-weak-best-response or forward induction, or using a combination of these techniques. Sequentially stable outcomes are extensive-form stable when they are unique. Our results on signaling games illustrate the strength of sequential stability. Sequentially stable outcomes pass most of the commonly used selection criteria, and they coincide with stable outcomes when they are unique or when payoffs are generic.

Several questions that are not addressed in our analysis may constitute avenues for future research. First, an axiomatic characterization of sequential stability would be desirable.[28] Second, it may be interesting to investigate which classes of games beyond signaling games feature generic equivalence between stability and sequential stability. Finally, studying strengthenings of sequential stability that maintain the existence of outcomes may help ensure uniqueness.

[28]This kind of characterization has proven to be elusive for stable sets. See Govindan and Wilson (2012) for a characterization of stable equilibria (as defined in Mertens, 1989) in two-player games with generic payoffs.

References

BAGWELL, K. (1990): "Informational product differentiation as a barrier to entry," *International Journal of Industrial Organization*, 8, 207–223.

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

BLUME, L. E. AND W. R. ZAME (1994): "The algebraic geometry of perfect and sequential equilibrium," *Econometrica: Journal of the Econometric Society*, 62, 783–794.

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.

CRAMTON, P. C. (1985): *Sequential bargaining mechanisms*, Cambridge University Press, Cambridge, England.

DILMÉ, F. (2023): "Lexicographic numbers in extensive form games," Mimeo.

FUDENBERG, D., D. M. KREPS, AND D. K. LEVINE (1988): "On the robustness of equilibrium refinements," *Journal of Economic Theory*, 44, 354–380.

FUDENBERG, D. AND J. TIROLE (1991a): *Game theory*, MIT press.

——— (1991b): "Perfect Bayesian equilibrium and sequential equilibrium," *Journal of Economic Theory*, 53, 236–260.

GOVINDAN, S. (1996): "A subgame property of stable equilibria," *Mathematics of Operations Research*, 21, 991–999.

GOVINDAN, S. AND R. WILSON (2012): "Axiomatic Equilibrium Selection for Generic Two-Player Games," *Econometrica: Journal of the Econometric Society*, 80, 1639–1699.

HARRINGTON, J. E. (1993): "The impact of reelection pressures on the fulfillment of campaign promises," *Games and Economic Behavior*, 5, 71–97.

JACKSON, M. O., T. RODRIGUEZ-BARRAQUER, AND X. TAN (2012): "Epsilon-equilibria of perturbed games," *Games and Economic Behavior*, 75, 198–216.

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.

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

Mertens, J.-F. (1989): "Stable Equilibria—A Reformulation: Part I. Definition and Basic Properties," *Mathematics of Operations Research*, 14, 575–625.

Myerson, R. B. (1978): "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, 7, 73–80.

Myerson, R. B. and P. J. Reny (2020): "Perfect Conditional ε-Equilibria of Multi-Stage Games With Infinite Sets of Signals and Actions," *Econometrica: Journal of the Econometric Society*, 88, 495–531.

Nash, J. (1951): "Non-cooperative games," *Annals of mathematics*, 54, 286–295.

Okada, A. (1981): "On stability of perfect equilibrium points," *International Journal of Game Theory*, 10, 67–73.

Osborne, M. J. and A. Rubinstein (1994): *A course in game theory*, MIT press.

Radner, R. (1980): "Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives," *Journal of Economic Theory*, 22, 136–154.

Riley, J. G. (1979): "Informational equilibrium," *Econometrica: Journal of the Econometric Society*, 47, 331–359.

Rubinstein, A. (1985): "A bargaining model with incomplete information about time preferences," *Econometrica: Journal of the Econometric Society*, 53, 1151–1172.

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.

Takahashi, S. and O. Tercieux (2020): "Robust equilibrium outcomes in sequential games under almost common certainty of payoffs," *Journal of Economic Theory*, 188, 105068.

van Damme, E. (1991): *Stability and perfection of Nash equilibria*, vol. 339, Springer.

A Proofs of the results

A.1 A useful result

Before proceeding to the proofs of the results, we state and prove a result that will be useful when proving results on sequentially stable outcomes. It establishes that sequential stability can be equivalently defined in an apparently weaker form than in our Definition 3.1.

Lemma A.1. *An outcome $\omega \in \Omega$ is sequentially stable if and only if for any vanishing tremble $(\eta_n)_n$ there is a strictly increasing sequence $(j_n \in \mathbb{N})_n$ and two sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_{j_n}))_n$ such that $(\omega^{\sigma_n})_n$ converges to ω.*

Proof. The "if" direction is obvious: when ω is sequentially stable, the result holds by setting $j_n := n$ for all $n \in \mathbb{N}$. Assume then that ω is such that for any vanishing tremble $(\eta_n)_n$ there is a strictly increasing sequence $(j_n \in \mathbb{N})_n$ and two sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_{j_n}))_n$ such that $(\omega^{\sigma_n})_n$ converges to ω. Fix some vanishing tremble $(\eta_n)_n$, and assume for the sake of contradiction that there is no pair of sequences $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ such that $(\omega^{\sigma_n})_n$ converges to ω. Then, there must be some $\varepsilon, \varepsilon' > 0$ and a strictly increasing sequence $(j_n \in \mathbb{N})_n$ such that $d(\omega, \Omega^*_\varepsilon(\eta_{j_n})) \geq \varepsilon'$ for all n, where $\Omega^*_\varepsilon(\eta_{j_n})$ is the set of perfect ε-equilibria of $G(\eta_{j_n})$ and

$$d(\omega, \Omega^*_\varepsilon(\eta_{j_n})) := \inf_{\omega' \in \Omega^*_\varepsilon(\eta_{j_n})} d(\omega, \omega')$$

(recall that, as explained in Footnote 7, we use the sup distance between outcomes; that is, for any pair of outcomes ω and ω', $d(\omega, \omega') := \max_{t \in T} |\omega(t) - \omega'(t)|$). This contradicts our original assumed property of ω, since the vanishing tremble $(\eta_{j_n})_n$ does not have a subsequence and corresponding sequences of epsilons and perfect epsilon-equilibria with outcomes converging to ω. □

A.2 Proofs of the results in Section 3

Proof of Proposition 3.1

Proof. **"Only if" part:** Assume ω is sequentially stable. Fix some $\varepsilon, \varepsilon' > 0$ and assume, for the sake of contradiction, that there is no $\delta_{\varepsilon,\varepsilon'} > 0$ such that $G(\eta)$ has a perfect ε-equilibrium with outcome ε'-close to ω for all $\|\eta\| < \delta_{\varepsilon,\varepsilon'}$. Then, there exists a vanishing tremble $(\eta_n)_n$ such that there is no $(\sigma_n \in \Sigma^*_\varepsilon(\eta_n))_n$ such that ω^{σ_n} is closer to ω than ε'. This contradicts that ω is sequentially stable.

"If" part: Assume that for all $\varepsilon, \varepsilon' > 0$ there is some $\delta_{\varepsilon,\varepsilon'} > 0$ such that, if $\|\eta\| < \delta_{\varepsilon,\varepsilon'}$, then $G(\eta)$ has a perfect ε-equilibrium with outcome at a distance lower than ε' from ω. Take a vanishing tremble

$(\eta_n)_n$. Let $\overline{u} := \max_{i\in N}(\max_{t\in T} u_i(t) - \min_{t\in T} u_i(t))$. Fix some sequence $(\hat{\varepsilon}_n)_n$ strictly decreasing towards 0 with $\hat{\varepsilon}_0 = \overline{u}$, and recursively define ε_n as follows:

1. We define $n_0 := 0$ and $\varepsilon_0 := \hat{\varepsilon}_0$.
2. For all $k \geq 1$ we let $n_k := \min\{n > n_{k-1} \mid \|\eta_{n'}\| < \delta_{\hat{\varepsilon}_k, \hat{\varepsilon}_k}$ for all $n' > n\}$. We let $\varepsilon_n := \hat{\varepsilon}_{k-1}$ for all $n = n_{k-1}+1, \dots, n_k$.

It is clear that $(\varepsilon_n)_n \to 0$. Note that, for each k, $\|\eta_n\| < \delta_{\hat{\varepsilon}_{k-1}, \hat{\varepsilon}_{k-1}}$ for all $n \in \{n_{k-1}+1, \dots, n_k\}$. Hence, there exists a sequence $(\sigma_n)_n$ where, for each k and $n \in \{n_{k-1}+1, \dots, n_k\}$, σ_n is a perfect $\hat{\varepsilon}_{k-1}$-equilibrium of $G(\eta_n)$ with outcome $\hat{\varepsilon}_{k-1}$-close to ω. Hence, since $\hat{\varepsilon}_{k-1} = \varepsilon_n$ for all $n \in \{n_{k-1}+1, \dots, n_k\}$, we have that $\omega^{\sigma_n} \to \omega$, and so ω is sequentially stable. □

Proof of Proposition 3.2

Proof. **"Only if" part:** Assume (σ, μ) is a sequential equilibrium supported by a fully-mixed sequence of strategy profiles $(\sigma_n)_n$. Let $A_* := \{a \in A | \sigma(a) > 0\}$, and define

$$\eta_n(a) := \begin{cases} \sigma_n(a) & \text{if } a \notin A_*, \\ 1/n & \text{otherwise.} \end{cases}$$

It is clear that $\eta_n(a) \to 0$ for all $a \in A$. We define

$$\varepsilon_n := 1/n + \max_{a\in A_*} \underbrace{\max_{a' \in A^{I^a}} (u(a'|\sigma_n) - u(a|\sigma_n))}_{(*)} .$$

Note that the term $(*)$ is non-negative (because $a \in A^{I^a}$). Note that, given our definitionof ε_n, we have that $u(a|\sigma_n) \geq u(a'|\sigma_n) - \varepsilon_n$ for all $a \in A_*$, $a' \in A^{I^a}$ and $n \in \mathbb{N}$, hence $\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n)$. Also, for all $a \in A_*$, we have that $(*)$ tends to 0 as $n \to +\infty$, since $u(a'|\sigma_n) \to u(a'|\sigma, \mu)$ for all a' and also $u(a|\sigma, \mu) = \max_{a' \in A^{I^a}} u(a'|\sigma, \mu)$ by sequential rationality. It is then clear that $\varepsilon_n \to 0$.

"If" part: We now fix some ω and assume that there exists a tremble sequence $(\eta_n)_n \to 0$, a sequence $(\varepsilon_n)_n \to 0$, and a sequence $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ such that $\omega^{\sigma_n} \to \omega$. Let $(k_n)_n$ be strictly increasing and satisfy that $(\sigma_{k_n})_n$ converges to some σ and has a corresponding sequence of belief systems converging to some μ. Then, take $a \in A$ such that $\sigma(a) > 0$ and some $a' \in A^{I^a}$. We then have that, since $u(a|\sigma_n) \geq u(a'|\sigma_n) - \varepsilon_n$ for all n high enough (because $\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n)$), we have $u(a|\sigma, \mu) \geq u(a'|\sigma, \mu)$. It is then clear that σ is sequentially rational under (σ, μ), which is supported by $(\sigma_{k_n})_n$, so (σ, μ) is a sequential equilibrium with outocme ω. □

Proof of Corollary 3.1

Proof. The proof is immediate from the arguments preceding the result. □

Proof of Proposition 3.3

Proof. The proof is immediate from the arguments in the main text. □

Proof of Proposition 3.4

Proof. Let $\hat{G}$ be the agent extensive form of G.[29] Let $\hat{N}$ and $\hat{u}$ denote the set of players and their payoff functions in $\hat{G}$, respectively. Note that $\hat{G}$ has the same set of sequentially stable outcomes as G. For each given payoff function $\tilde{u} \equiv (\tilde{u}_i : T \to \mathbb{R})_{i \in \hat{N}}$, we let $\hat{G}(\tilde{u})$ be the agent extensive form of the game defined in Section 2 with payoff function given by $\tilde{u}$ instead of $\hat{u}$. Let $(\hat{u}_m)_m$ be a sequence of payoff functions converging to $\hat{u}$ such that, for each m, $G(\hat{u}_m)$ has a stable outcome denoted ω_m, which by Proposition 3.3 is also sequentially stable.[30] Note that, since a stable outcome exists for generic payoff functions (by Kohlberg and Mertens, 1986), a sequence $(\hat{u}_m)_m$ with the previous properties exists. We can assume without loss of generality for our argument that $(\omega_m)_m$ converges to some outcome ω. Then, it follows from the following lemma that ω is sequentially stable.

Lemma A.2. *Let $(u_m)_m$ be a sequence of payoff functions converging to u. Let $(\omega_m)_m \to \omega$ be such that each ω_m is sequentially stable in $G(u_m)$. Then, ω is a sequentially stable outcome of G.*

Proof. Fix a vanishing tremble $(\eta_n)_n$. For each $m \in \mathbb{N}$, let $(\varepsilon_{m,n})_n \to 0$ and $(\sigma_{m,n} \in \Sigma^*_{\varepsilon_{m,n}}(\eta_n, u_m))_n$ be such that $\omega^{\sigma_{m,n}} \to \omega_m$ as $n \to \infty$, which exist by the assumption that ω_m is sequentially stable in $G(u_m)$. Note that for each $a \in A$ and $m, n \in \mathbb{N}$ with $\sigma_{m,n}(a) > \eta_n(a)$, we have

$$u_m(a|\sigma_{m,n}) \geq u_m(a'|\sigma_{m,n}) - \varepsilon_{m,n} \text{ for all } a' \in A^{I^a}.$$

Hence,

$$u(a|\sigma_{m,n}) \geq u(a'|\sigma_{m,n}) - 2\,d(u, u_m) - \varepsilon_{m,n} \text{ for all } a' \in A^{I^a}.$$

---

[29] That is, $\hat{G}$ coincides with G except that $\hat{N}$ is bigger than N, each information set is associated to a different player, and each player associated to a given information set has the same payoff for each terminal history as the player associated to this information set in G.

[30] Recall that, as explained in Footnote 7, we use the sup distance between payoff functions; that is, for any pair of payoff functions u and u', $d(u, u') := \max_{i \in N} \max_{t \in T} |u_i(t) - u'_i(t)|$.

Let $(n_m)_m$ be a sequence of indexes such that $\varepsilon_{m,n_m} \to 0$ and $\omega^{\sigma_{m,n_m}} \to \omega$ as $m \to \infty$, which exists by a standard diagonal argument.[31] Then, defining $\varepsilon_m := 2\,d(u, u_m) + \varepsilon_{m,n_m}$, we have that each $\sigma_m := \sigma_{m,n_m}$ is a perfect ε_m-equilibrium of $G(\eta_{n_m})$. Then, since $\varepsilon_m \to 0$ as $m \to \infty$, and since the argument holds for any vanishing tremble $(\eta_n)_n$, Lemma A.1 implies that ω is sequentially stable. □

□

A.3 Proofs of the results in Section 4

Proof of Proposition 4.1

Proof. Let ω be a sequentially stable outcome. Let $\hat{a} \in A^I$ be an action that is *not* sequentially optimal in any sequential equilibrium with outcome ω (hence $\hat{a}$ is not played under ω, either because $I^{\hat{a}}$ is off-path or because $I^{\hat{a}}$ is on path but $\hat{a}$ is chosen with probability zero). Let G' denote the game where $\hat{a}$ (and all consecutive histories) is eliminated, and $A' \subset A \setminus \{\hat{a}\}$ be its set of actions. Let $(\eta'_n)_n$ be a vanishing tremble of G', and let $\underline{\eta}'_n := \min\{\eta'_n(a') | a' \in A'\}$. Define the vanishing tremble $(\eta_n)_n$ as follows:

$$\eta_n(a) := \begin{cases} \eta'_n(a) & \text{if } a \in A', \\ (\underline{\eta}'_n)^{|A|} & \text{otherwise,} \end{cases}$$

for all $a \in A$, and note that $(\eta_n)_n$ is a vanishing tremble of G. Note also that, under the vanishing tremble $(\eta_n)_n$, any history not belonging to G' (i.e., with some $a \notin A'$) has a vanishing relative likelihood with respect to any history of G'. Let $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ be such that $\omega^{\sigma_n} \to \omega$ as $n \to +\infty$ (which exists since ω is sequentially stable). Taking a subsequence if necessary, assume that $(\sigma_n)_n$ supports some assessment (σ, μ), which by Proposition 3.2 is a sequential equilibrium, so (σ, μ) has outcome ω.[32] Note that, if n is large enough, it must be that $\sigma_n(\hat{a}) = \eta_n(\hat{a})$, since by assumption $\hat{a}$ is not sequentially optimal under (σ, μ), hence there is some $a' \in A^{\mathcal{I}^a}$ such that

$$\lim_{n\to\infty} u(\hat{a}|\sigma_n) = u(\hat{a}|\sigma, \mu) < u(\hat{a}'|\sigma, \mu) = \lim_{n\to\infty} u(\hat{a}'|\sigma_n)\ .$$

[31]Indeed, the diagonal argument sets $n_0 := 1$ and, for all $m > 0$, $n_m := \min\{n > n_{m-1} \,|\, \max\{\varepsilon_{m,n}, d(\omega^{\sigma_{m,n}}, \omega)\} < 1/m\}$.

[32]Recall that, by Lemma A.1, it is enough to prove that the property in Definition 3.1 holds for a subsequence of $(\eta_n)_n$.

Let $\hat{a}' \in A^{I^{\hat{a}}}$ be an action that is played with positive probability under σ. Define, for all $a' \in A'$,

$$\sigma'_n(a') := \begin{cases} \sigma_n(a') + \sigma_n(\hat{a}) & \text{if } a' = \hat{a}', \\ \sigma_n(a') & \text{if } a' \neq \hat{a}'. \end{cases}$$

Note that $\sigma'_n \in \Sigma'(\eta'_n)$ (i.e., belongs to the set of strategies of the perturbed game $G'(\eta'_n)$). We claim that there is some sequence $(\varepsilon'_n)_n \to 0$ such that $(\sigma'_n \in \Sigma'^*_{\varepsilon'_n}(\eta'_n))_n$. This follows from the fact that all information sets of G that contain both histories in G' and not in G', the relative weight of histories not in G' shrinks to 0 as n increases, since all of them have $\hat{a}$ as one of its elements. It then follows that, as n increases, all actions $a' \in A'$ with $\sigma(a') > 0$ are asymptotically sequentially optimal as $n \to \infty$.[33] □

Proof of Corollary 4.1

Proof. **Forward induction:** Let ω be sequentially stable, and I and a satisfy the conditions in the statement. It is then clear that a is not sequentially optimal under any sequential equilibrium with outcome ω. Hence, the result holds from applying NWBR.

Iterated strict equilibrium dominance: Let ω be sequentially stable and a satisfy the conditions in the statement. Since a is not sequentially optimal under any sequential equilibrium, it is not sequentially optimal under any sequential equilibrium with outcome ω. Hence, the result holds from applying NWBR. □

Proof of Proposition 4.2

Proof. We prove the equivalence between parts 1 and 3. Proving that part 1 is equivalent to part 2 is easy using Proposition 3.1.

Proof that $3 \Rightarrow 1$**:** Assume that for all $\varepsilon, \varepsilon' > 0$ there is some $\delta > 0$ with the property that, if $\|\eta\| < \delta$, then there is some u' with $\|u' - u\| < \varepsilon$ such that $G(\eta, u')$ has a Nash equilibrium with outcome ε'-close to ω. Fix some $\varepsilon, \varepsilon' > 0$, and let δ satisfy the aforementioned property. We then have that, if $\|\eta\| < \delta$, then $G(\eta)$ has a perfect ε-equilibrium with outcome ε'-close to ω. Applying Proposition 3.1, we then have that ω is sequentially stable.

---

[33] Even though we proved that a subsequence of $(\eta'_n)_n$ is such that there are $(\varepsilon'_n)_n \to 0$ and $(\sigma'_n \in \Sigma^*_{\varepsilon'_n}(\eta'_n|G'))_n$ with $\omega^{\sigma_n} \to \omega$, Lemma A.1 ensures that this is enough to prove the sequential stability of ω in G'.

Proof that $1 \Rightarrow 3$**:** Let ω be a sequentially stable outcome. Fix some $\varepsilon, \varepsilon' > 0$. By Proposition 3.1 we have that there is some $\delta > 0$ such that, if $\|\eta\| < \delta$, then $G(\eta)$ has a perfect ε-equilibrium with outcome ε'-close to ω. We fix some η with $\|\eta\| < \delta$ and let $\sigma \in \Sigma^*_\varepsilon(\eta)$ be a perfect ε-equilibrium with outcome ε'-close to ω. We want to show that there is some u' with $\|u' - u\| < |\mathcal{I}|\,\varepsilon$ such that σ is a Nash equilibrium of $G(\eta, u')$.

We propose an algorithm that will change the payoff function from u to some u' with the desired property. It will do so by changing the payoffs of the terminal histories in so that ε-optimal actions under u will become exactly optimal under u'. To do so, recall that $T^a \subset T$ is the set of terminal histories containing $a \in A$. We denote the information sets $\mathcal{I} := \{I_1, ..., I_{|\mathcal{I}|}\}$. We define $\hat{u}^j$ recursively from $j = 1$ to $|\mathcal{I}|$, and we initialize $\hat{u}^0 := u$. As we shall see, in each step j, the expected continuation payoff difference between two actions played at any information set different from I_j remains unchanged. For each $j = 1, ..., |\mathcal{I}|$, we proceed as follows:

1. Define $\overline{u}^j_{\iota(I_j)} := \max_{a \in A^{I_j}} \hat{u}^{j-1}(a|\sigma)$. Let $A^{I_j}_*$ be the set of actions $a \in A^{I_j}$ such that $u(a) \geq \overline{u}_{\iota(I_j)} - \varepsilon$. Note that $a \in A^{I_j}$ is such that $\sigma(a) > \eta(a)$ only if $a \in A^{I_j}_*$.
2. We define $\hat{u}^j$ as a payoff function assigning to each $i \in N$ and $t \in T$ the value $\hat{u}^{j-1}_i(t)$, except for the value assigned to player $\iota(I_j)$ at terminal histories $t \in T^a$ for some $a \in A^{I_j}_*$, where

$$\hat{u}^j_{\iota(I_j)}(t) := \hat{u}^{j-1}_{\iota(I_j)}(t) + \underbrace{\overline{u}^j_{\iota(I_j)} - \hat{u}^{j-1}(a|\sigma)}_{\geq 0} - K(I_j)\ , \tag{A.1}$$

   where $K(I_j)$ is chosen such that

$$\mathbb{E}[\hat{u}^j_{\iota(I_j)}(t)|\sigma, I_j] = \mathbb{E}[\hat{u}^{j-1}_{\iota(I_j)}(t)|\sigma, I_j]\ . \tag{A.2}$$

   Note that $K(I_j) \in [0, \varepsilon)$. Note also that, under $\hat{u}^j$, player $\iota(I_j)$'s continuation payoff is the same for all $a \in A^{I_j}_*$, (and equal to $\hat{u}^j(a|\sigma) = \overline{u}^j_{\iota(I_j)} - K(I_j)$), which is weakly higher than $\hat{u}^j(a'|\sigma)$ for all $a' \in A^{I_j}$. This guarantees that player $\iota(I_j)$ plays a best response to σ at I_j under $\hat{u}^j$.
3. Note that, for each $a \in A^{I_j}$, we have that

$$\hat{u}^j_{\iota(I_j)}(t) - \hat{u}^j_{\iota(I_j)}(t') = \hat{u}^{j-1}_{\iota(I_j)}(t) - \hat{u}^{j-1}_{\iota(I_j)}(t')$$

   for all $t, t' \in T^a$, hence payoff difference from choosing an action instead of another at an information set of player $\iota(I_j)$ that follows I_j remains the same. Condition (A.2) guarantees that the continuation payoff player $\iota(I_j)$ obtains at information set I_j is the same under $\hat{u}^{j-1}$ and under $\hat{u}^j$, hence her incentives in one of her information sets that preceeds I_j remains the same. Note finally that $\|\hat{u}^j - \hat{u}^{j-1}\| < \varepsilon$.

Using the triangle inequality, we have that

$$\|u-\hat{u}^{|\mathcal{I}|}\| \leq \sum_{j=1}^{|\mathcal{I}|} \|\hat{u}^j-\hat{u}^{j-1}\| < |\mathcal{I}|\,\varepsilon\ .$$

Since, as we argued, each player plays a best response to σ in each information set under $\hat{u}^{|\mathcal{I}|}$, we have that the desired u' is $\hat{u}^{|\mathcal{I}|}$. □

Proof of Proposition 4.3

Proof. **Proof of part 1:** Let G' be a subgame of G originated at an information set denoted I'. Assume that G' has a unique sequential outcome ω', hence ω' is a sequentially stable outcome of G' (by Corollary 3.1 and Proposition 3.4). Let $\hat{G}$ be the game obtained by replacing G' by ω' (recall Footnote 19). Let A' be the set of actions of G', $T'_{\omega'}$ be the support of ω', and $\hat{A}$ be $A\backslash A'$. We want to prove that G and $\hat{G}$ have the same set of stable outcomes. We divide the proof into two subparts.

1. Let ω be a sequentially stable outcome of G. Take some vanishing tremble $(\hat{\eta}_n)_n$ in $\hat{G}$ and some vanishing tremble $(\eta'_n)_n$ in G'. Let $(\eta_n)_n$ be defined as

$$\eta_n(a) := \begin{cases} \eta'_n(a) & \text{if } a \in A', \\ \hat{\eta}_n(a) & \text{if } a \in \hat{A}, \end{cases}$$

for all $a \in A$. Let $(\varepsilon_n)_n \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n))_n$ be such that $\omega^{\sigma_n} \to \omega$ (which exist because ω is stable in G). By Proposition 3.2 and the fact that ω' is the unique sequential outcome in G', we have that the conditional distribution of ω^{σ_n} on T' converges to ω' as $n \to \infty$. Let $\hat{\sigma}_n$ be defined as $\hat{\sigma}_n(a) := \sigma_n(a)$ for all $a \in A$ and $\hat{\sigma}_n(t') := \omega'(t')$ for all terminal histories $t' \in T'_{\omega'}$ (note that, in G', nature plays each t' with probability $\omega'(t')$ at information set I'). It is clear that $\omega^{\hat{\sigma}_n} \to \omega$. Also, it is easy to see that there exists some $(\hat{\varepsilon}_n)_n \to 0$ such that $\hat{\sigma}_n \in \hat{\Sigma}^*_{\hat{\varepsilon}_n}(\hat{\eta}_n)$ for all n. Hence, ω is sequentially stable in $\hat{G}$.

2. Let $\hat{\omega}$ be a sequentially stable outcome of $\hat{G}$. Take some vanishing tremble $(\eta_n)_n$ in G and let $(\eta'_n)_n$ be its restriction to G'. Let $(\hat{\eta}_n)_n$ be a vanishing tremble in $\hat{G}$ satisfying that $\hat{\eta}_n(a) := \eta_n(a)$ for all $a \in \hat{A}$ and $\hat{\eta}_n(t') \leq \omega'(t')$ for all $t' \in T'_{\omega'}$. Let $(\hat{\varepsilon}_n)_n \to 0$ and $(\hat{\sigma}_n \in \hat{\Sigma}^*_{\varepsilon'_n}(\hat{\eta}_n))_n$ be such that $\omega^{\hat{\sigma}_n} \to \hat{\omega}$ (which exist because $\hat{\omega}$ is sequentially stable in $\hat{G}$). Let $(\varepsilon'_n)_n \to 0$ and $(\sigma'_n \in \Sigma'^*_{\varepsilon'_n}(\eta'_n))_n$ be such that $\omega^{\sigma'_n} \to \omega'$ (which exist by Proposition 3.2 and the fact that ω' is the unique sequential outcome in G'). Let σ_n be defined as $\sigma_n(a) := \hat{\sigma}_n(a)$ for all $a \in \hat{A}$ and $\sigma_n(a) := \sigma'_n(a)$ for all $a \in A'$. It is then clear that $\omega^{\sigma_n} \to \hat{\omega}$. Again, it is easy to see that there exists $(\varepsilon_n)_n \to 0$ such that $\sigma_n \in \Sigma^*_{\varepsilon_n}(\eta_n)$ for all n. Hence, $\hat{\omega}$ is sequentially stable in G.

Proof of part 2: The proof is similar to that of the second case in part 1, and hence omitted.

Proof of part 3: Let G' be a subgame of G played with positive probability under ω, and let T' be the set of terminal histories of G'.[34] Assume, for contradiction, that the restriction of ω to G', $\omega' := \omega|_{T'}$, is not a sequentially stable outcome of G', and let $(\eta'_n)_n$ be a tremble sequence of G' destroying it (i.e., such that there are no corresponding sequences $(\varepsilon'_n)_n$ and $(\sigma'_n)_n$ with the properties in Definition 3.1). Then, it is clear that any perturbation η_n that coincides with η'_n when restricted to G' destroys ω, contradicting that ω is a sequentially stable outcome of G. □

Proof of Proposition 4.4

Proof. Let $I, I' \in \mathcal{I}$ be such that $I' = I \times A^I$. For each terminal history $(a_1, ..., a_J) \in T$, define

$$\mathcal{T}(a_1, ..., a_J) := \begin{cases} (a_1, ..., a_{j+1}, a_j, ..., a_J) & \text{if } a_j \in A^I \text{ for some } j, \\ (a_1, ..., a_J) & \text{otherwise.} \end{cases}$$

Let G' be a game obtained from G by replacing T by $\mathcal{T}(T)$, and also replacing H, $\mathcal{I}$, and ι accordingly (note that the set of actions does not change). Let $u' := u \circ \mathcal{T}$ be the payoff function in G'.

We now fix some sequentially stable outcome ω of G, and we will show that the outcome analogous to ω in G', denoted $\omega' := \omega \circ \mathcal{T}^{-1}$, is also sequentially stable. To see this, fix some vanishing tremble $(\eta_n)_n$, and let $(\varepsilon_n)_n$ and $(\sigma_n)_n$ satisfy the conditions in Definition 3.1, with $(\omega^{\sigma_n})_n \to \omega$ (which exist since ω is sequentially stable). We argue that $(\eta_n)_n$, $(\varepsilon_n)_n$ and $(\sigma_n)_n$ also satisfy the conditions in Definition 3.1 in G' and the limit outcome is ω'.[35] Then, we have that player $\iota(I^a)$'s payoff from playing a in G under σ_n is

$$u(a|\sigma_n) = \frac{\sum_{t \in T^a} \Pr^{\sigma_n}(t)\, u_{\iota(I^a)}(t)}{\Pr^{\sigma_n}(I^a)\, \sigma_n(a)} = \frac{\sum_{t \in T^a} \Pr^{\sigma_n}(\mathcal{T}(t))\, u'_{\iota(I^a)}(\mathcal{T}(t))}{\Pr^{\sigma_n}(\mathcal{T}(I^a))\, \sigma_n(a)} = u'(a|\sigma_n) ,$$

where the second equality follows because of $\Pr^{\sigma_n}(\mathcal{T}(t)) = \Pr^{\sigma_n}(t)$ (since t and $\mathcal{T}(t)$ contain the same actions), $u'_{\iota(I^a)}(\mathcal{T}(t)) = u_{\iota(I^a)}(t)$ (by definition of u') and $\Pr^{\sigma_n}(\mathcal{T}(I^a)) = \Pr^{\sigma_n}(I^a)$ (because $\mathcal{T}$ applied to all terminal histories that follow I^a in G equals the set of all terminal histories that follow $\mathcal{T}(I^a)$ in G'). It is then clear that $(\eta_n)_n$, $(\varepsilon_n)_n$ and $(\sigma_n)_n$ also satisfy the conditions in Definition 3.1 in G', hence ω' is a sequentially stable outcome of G'. □

[34] We use the standard definition of subgame (e.g., from Osborne and Rubinstein, 1994).

[35] Mote that since G' has the same set of actions as G and since, for each $a \in A$, the set of available actions at the information set where a is available is the same in both G and G', we have that $(\eta_n)_n$ is also a vanishing tremble of G', and σ_n is a perfect ε_n-equilibrium of $G'(\eta_n)$ for each n.

Proof of Proposition 4.5

Proof. We assume that ω is the unique sequentially stable outcome of G and, for the sake of contradiction, assume that it is not extensive-form stable. Note that there is no extensive-form stable outcome of G different from ω, since otherwise, such an outcome would also be sequentially stable (by Proposition 3.3), contradicting the assumption that ω is the unique sequentially stable outcome. Hence, it must be that ω is *not* extensive-form stable.

Let $\hat{G}$ be the agent extensive form of G and let $\hat{u}$ be its corresponding payoff function. The previous assumptions imply that ω is the unique sequentially stable outcome of $\hat{G}$ and that $\hat{G}$ has no stable outcome. Let $(\hat{\eta}_{\hat{n}})_{\hat{n}}$ be a tremble such that there is no sequence of indexes $(\hat{n}_j)_j$ such that there is some sequence $(\hat{\sigma}_j \in \Sigma_0^*(\hat{\eta}_{\hat{n}_j}))_j$ with outcomes converging to ω (where $(\hat{\eta}_{\hat{n}})_{\hat{n}}$ exists since ω is not stable in $\hat{G}$). By Proposition 3.4, each $\hat{G}(\hat{\eta}_{\hat{n}})$ has at least one sequentially stable outcome $\omega_{\hat{n}}$. We let $(\hat{n}_j)_j$ be a sequence of indexes such that $\omega_{\hat{n}_j}$ converges to some ω'. Since, for each j, $\omega^{\hat{\sigma}_j} = \omega_{\hat{n}_j}$ for some $\hat{\sigma}_j \in \Sigma_0^*(\hat{\eta}_{\hat{n}_j})$, the previous assumption on $(\hat{\eta}_{\hat{n}})_{\hat{n}}$ implies that $\omega' \neq \omega$. We then reach a contradiction by proving that ω' is sequentially stable. This follows from the following result, which is analogous to Lemma A.2.

Lemma A.3. *Let $(\hat{\eta}_{\hat{n}})_{\hat{n}}$ be a vanishing tremble. Let $(\omega_{\hat{n}})_{\hat{n}} \to \omega$ be such that each $\omega_{\hat{n}}$ is sequentially stable in $G(\hat{\eta}_{\hat{n}})$. Then, ω is a sequentially stable outcome of G.*

Proof. The proof is similar to that of Lemma A.2, and left to the reader. □

□

A.4 Proofs of the results in Section 5

Proof of Proposition 5.1

Proof. In this proof, we will use the following notation. For a given strategy profile $\sigma \in \Sigma^{\text{sig}}$, we will use $\sigma(m|\theta)$ and $\sigma(r|m)$ to indicate the probability with which the sender chooses m after θ and the probability with which the receiver chooses r after m, respectively.

"Only if" direction. Assume ω is a sequentially stable outcome. Let m be a message unsent under ω. Take a probability distribution μ_m over Θ_m and a vanishing tremble $(\eta_n)_n$ such that

$$\lim_{n\to\infty} \frac{\pi(\theta)\,\eta_n(m|\theta)}{\sum_{\theta'\in\Theta_m} \pi(\theta')\,\eta_n(m|\theta')} = \mu_m(\theta)$$

and such that $\eta_n(m|\theta')/\eta_n(m|\theta)=\mu_m(\theta')/\mu_m(\theta)$ for all $\theta,\theta'\in\Theta_m$ with $\theta\in\text{supp}(\mu_m)$.[36] Let $(\varepsilon_n)_n$ and $(\sigma_n)_n$ be two sequences satisfying that $\varepsilon_n\to 0$, $\sigma_n\in\Sigma^*_{\varepsilon_n}(\eta_n)$ for all n, and $\omega^{\sigma_n}\to\omega$ (they exist because ω is a sequentially stable outcome). Taking a subsequence if necessary, assume that $(\sigma_n)_n$ supports some sequential equilibrium (σ,μ'') (with outcome ω). Note that $u_\theta(m|\sigma)\le u_\theta(\omega)$ for all $\theta\in\Theta_m$. There are two cases:

1. Assume first that $u_\theta(m|\sigma)=u_\theta(\omega)$ for all θ with $\mu''(\theta|m)>0$. Then, the result holds for $\alpha:=0$, $\mu'_m(\cdot):=\mu''(\cdot|m)$, and $\rho:=\sigma(\cdot|m)$.

2. Assume now that $u_\theta(m|\sigma)<u_\theta(\omega)$ for some θ with $\mu''(\theta|m)>0$. Then, there is some $\overline{n}$ such that $\sigma_n(m|\theta)=\eta_n(m|\theta)$ for all $n>\overline{n}$. Note further that it must be that $\theta\in\text{supp}(\mu_m)$, since by the definition of $(\eta_n)_n$, we have that $\eta_n(m|\theta')/\eta_n(m|\theta'')=0$ whenever $\theta'\notin\text{supp}(\mu_m)$ and $\theta''\in\text{supp}(\mu_m)$. There are then two cases:

    (a) If $\mu_m(\theta)=\mu''(\theta|m)$, then it must be that $\mu_m(\theta')=\mu''(\theta'|m)$ for all $\theta'\in\Theta_m$.[37] Hence, the result holds for $\alpha:=1$, $\mu'_m(\cdot):=\mu''(\cdot|m)$, and $\rho:=\sigma(\cdot|m)$.

    (b) If $\mu_m(\theta)\neq\mu''(\theta|m)$ then it must be that $\mu_m(\theta)<\mu''(\theta|m)$, since

    $$\begin{aligned}\mu''(\theta|m) &= \lim_{n\to\infty}\frac{\pi(\theta)\,\eta_n(m|\theta)}{\sum_{\theta'\in\Theta_m}\pi(\theta')\,\sigma_n(m|\theta')}\\ &\le \lim_{n\to\infty}\frac{\pi(\theta)\,\eta_n(m|\theta)}{\sum_{\theta'\in\Theta_m}\pi(\theta')\,\eta_n(m|\theta')}=\mu_m(\theta)\,.\end{aligned}$$

    Define $\alpha:=1-\mu_m(\theta)/\mu''(\theta|m)\in(0,1]$, so $\mu''(\theta|m)=(1-\alpha)\,\mu_m(\theta)$. Note that, for any other θ' such that $u_{\theta'}(m|\sigma)<u_{\theta'}(\omega)$ it must be that $\mu''(\theta'|m)=(1-\alpha)\,\mu_m(\theta')$, since $\sigma_n(\theta')=\eta_n(\theta')$ for n large enough in this case. We then have that the result holds for the obtained value of α, for $\mu'_m(\theta):=(\mu''(\theta|m)-(1-\alpha)\,\mu_m(\theta))/\alpha$, and for $\rho:=\sigma(\cdot|m)$.

"If" direction. Assume ω satisfies the condition in the statement of Proposition 5.1. We fix a vanishing tremble $(\eta_n)_n$. We will construct a strictly increasing sequence $(j_n)_n$ and a sequence $(\sigma_{j_n})_n$ such that $\sigma_{j_n}\in\Sigma^*_{\varepsilon_{j_n}}(\eta_n)$ for all n for some sequence $(\varepsilon_{j_n})_n\to 0$ and $\omega^{\sigma_{j_n}}\to\omega$ as $n\to\infty$; hence, the sequential stability of ω will follow from Lemma A.1.

We denote the messages which are off path of ω as $M_0:=\{m^1,\dots,m^{|M_0|}\}$ (we assume that $|M_0|\ge 1$ since the result is trivial otherwise). We first construct $\sigma_n(m|\theta)$ and $\sigma_n(r|m)$ for all $m\in M_0$,

[36] For example, $\eta_n(m|\theta):=\pi(\theta)^{-1}\,\mu(\theta)\,n^{-1}$ for all $\theta\in\text{supp}(\mu)$ and $\eta_n(m|\theta):=n^{-2}$ for all $\theta\notin\text{supp}(\mu)$.

[37] Indeed, because $\sigma_n(m|\theta)=\eta_n(m|\theta)$ for large n, $\mu_m(\theta)=\mu''(\theta|m)$ only if $\lim_{n\to\infty}\sigma_n(m|\theta')/\eta_n(m|\theta')=1$ for all $\theta'\in\text{supp}(\mu'')$ and $\lim_{n\to\infty}\sigma_n(m|\theta')/\eta_n(m|\theta)=0$ for all $\theta'\notin\text{supp}(\mu'')$.

$\theta \in \Theta_m$, and $r \in R_m$. We do it by first proceeding recursively over the set of messages which are off-path under ω, and then we will define the values for on-path messages. We begin with $k=1$ and $(j_n^0) := (n)_n$. Then, for each $k=1,\dots,|M_0|$, we proceed as follows:

1. We let $(j_n^k)_n$ be a strictly increasing subsequence of $(j_n^{k-1})_n$ such that
$$\mu_{m^k}(\theta) := \lim_{n\to\infty} \frac{\pi(\theta)\,\eta_{j_n^k}(m^k|\theta)}{\sum_{\theta'\in\Theta_{m^k}} \pi(\theta')\,\eta_{j_n^k}(m^k|\theta')}$$
is well defined for all $\theta \in \Theta_{m^k}$.

2. Let μ'_{m^k}, α, and ρ be the ones determined by the statement for μ_{m^k} and m^k.

3. There are two cases:

    (a) If $\alpha=1$ then we set $\sigma_n(m^k|\theta) := \eta_n(m^k|\theta)$ for all $\theta\in\Theta_{m^k}$ and $(j_n^k)_n := (\hat{j}_n^k)_n$.

    (b) If $\alpha\neq 1$ then let $K_n := \sum_{\theta\in\Theta_{m^k}} \mu_{m^k}(\theta)\,\eta_n(m^k|\theta)$. We then define, for each $\theta\in\Theta_{m^k}$,
    $$\sigma_n(m^k|\theta) := \eta_n(m^k|\theta) + \tfrac{\alpha}{1-\alpha}\,K_n\,\mu'_{m^k}(\theta)\ .$$
    Note that, for all $\theta\in\Theta_{m^k}$, we have
    $$\begin{aligned}\lim_{n\to\infty} \frac{\pi(\theta)\,\sigma_n(m^k|\theta)}{\sum_{\theta'\in\Theta(m^k)} \pi(\theta')\,\sigma_n(m^k|\theta')} &= \lim_{n\to\infty} \frac{\pi(\theta)\,(\eta_n(m^k|\theta) + \frac{\alpha}{1-\alpha}\,K_n\,\mu'_{m^k}(\theta))}{K_n + \frac{\alpha}{1-\alpha}\,K_n} \\ &= (1-\alpha)\,\mu_{m^k}(\theta) + \alpha\,\mu'_{m^k}(\theta)\ .\end{aligned}$$

4. We finally define $\sigma_n(r|m^k)$ as
$$\sigma_n(r|m^k) := \begin{cases} \eta_n(r|m^k) & \text{if } \rho(r)=0, \\ \rho(r)\big(1-\sum_{r'\notin \mathrm{supp}(\rho)} \eta_n(r'|m^k)\big) & \text{if } \rho(r)>0. \end{cases} \tag{A.3}$$
Note that, as $n\to\infty$, we have that $\sigma_n(m^k|\theta)\to\rho$.

For all messages m that occur on path under ω (that is, $m\notin M_0$), we define
$$\sigma_n(m|\theta) := \begin{cases} \eta_n(m|\theta) & \text{if } \omega(m|\theta)=0, \\ \omega(m|\theta)\big(1-\sum_{\theta'|\omega(m|\theta')=0} \eta_n(m|\theta')\big) & \text{if } \omega(m|\theta)>0, \end{cases}$$
where $\omega(m|\theta)$ is the probability that type θ sends m under ω, and also
$$\sigma_n(r|m) := \begin{cases} \eta_n(r|m) & \text{if } \omega(r|m)=0, \\ \omega(r|m)\big(1-\sum_{r'|\omega(r'|m)=0} \eta_n(r'|m)\big) & \text{if } \omega(r|m)>0, \end{cases}$$

where $\omega(r|m)$ is the probability that the receiver chooses r after m under ω (where m is an on-path message). It is not difficult to see that our construction (together with the properties of μ'_m, α, and ρ) guarantees that $\sigma_n \in \Sigma(\eta_n)$, that $\omega^{\sigma_{j_n}} \to \omega$, and that there is some sequence $(\varepsilon_n) \searrow 0$ such that $\sigma_{j_n} \in \Sigma^*_{\varepsilon_{j_n}}(\eta_{j_n})$ for all n (note that, by Lemma A.1, showing the convergence for a subsequence is enough to show sequential stability). $\square$

Proof of Corollary 5.1

Proof. As indicated in the main text, the proof is immediate from our Proposition 5.1 and Theorem 3 and Proposition 4 in Banks and Sobel (1987) and Cho and Kreps (1987), respectively. $\square$

Proof of Proposition 5.2

Proof. **Part 0: Notation.** To formally define a stable outcome, we need some notation regarding the normal form of the extensive game G defined in Section 2 (which is naturally extended to G^{sig}). For each player i, we let $\vec{A}^i := \prod_{I \in \iota^{-1}(i)} A^I$ denote her set of her *(normal-form) pure strategies* and $\vec{a} \equiv (\vec{a}^i)_{i \in N}$ be a generic pure strategy. We use $a \in \vec{a}^i$ to denote that $a \in A$ is one of the components of $\vec{a}^i \in \vec{A}^i$. We also let $\hat{\Sigma}^i := \Delta(\vec{A}^i)$ be player i's set of *(normal form) mixed strategies* and let $\hat{\sigma} \equiv (\hat{\sigma}^i)_{i \in N}$ be a generic mixed strategy, for each $i \in N \cup \{0\}$ (where nature plays according to the corresponding mixed strategy consistent with π). A *normal-form vanishing tremble* is a sequence $(\hat{\eta}_n)_n \equiv ((\hat{\eta}^i_n)_{i \in N})_n$, where $\hat{\eta}^i_n : \vec{A}^i \to (0,1]$ is such that $\hat{\eta}^i_n(\vec{a}^i) \to 0$ as $n \to \infty$ for all $\vec{a}^i \in \vec{A}^i$. Then, ω is *stable* if, for any normal-form vanishing tremble $(\hat{\eta}_n)_n$, there is a sequence $(\hat{\sigma}_n)_n$ such that (i) $\hat{\sigma}^i_n(\vec{a}^i) \geq \hat{\eta}^i_n(\vec{a}^i)$ for all $\vec{a}^i$ and n, (ii) $\hat{\sigma}^i_n(\vec{a}^i) > \hat{\eta}^i_n(\vec{a}^i)$ only if $\vec{a}^i$ is optimal, for all $\vec{a}^i$ and n, and (iii) $\omega^{\hat{\sigma}_n} \to \omega$ as $n \to \infty$.

Part 1: Proof that if ω is extensive-form stable in G^{sig}, then it is stable. Let ω be an extensive-form stable in G^{sig}. Let $(\hat{\eta}_n)_n$ be a normal-form vanishing tremble. Define the following extensive-form vanishing tremble for all a and n:

$$\eta_n(a) := \sum_{\vec{a} \ni a} \hat{\eta}^i_n(\vec{a}^i) \ .$$

We define $\overline{\eta}^i := \sum_{\vec{a}^i \in \vec{A}^i} \hat{\eta}^i_n(\vec{a}^i)$. Note that, for any information set I with $i = \iota(I)$, we have

$$\sum_{a \in A^I} \eta_n(a) = \sum_{a \in A^I} \sum_{\vec{a}^i \ni a} \hat{\eta}^i_n(\vec{a}^i) = \sum_{\vec{a}^i \in \vec{A}^i} \hat{\eta}^i_n(\vec{a}^i) = \overline{\eta}^i \ .$$

Since ω is extensive-form stable, there is a sequence of Nash equilibria $(\sigma_n \in \Sigma^*_0(\eta_n))_n$ with outcomes converging to ω. Taking a subsequence if necessary, assume $(\sigma_n)_n$ supports an assessment

(σ,μ). We let $\vec{A}^i_*$ be the set of action vectors $\vec{a}^i$ such that a is sequentially optimal under (σ,μ) for all $a\in\vec{a}^i$. We now define

$$\hat{\sigma}^i_n(\vec{a}^i):=\begin{cases}\hat{\eta}^i_n(\vec{a}^i) & \text{if } \sigma_n(a)=\eta_n(a) \text{ for some } a\in\vec{a}^i,\\ \frac{\prod_{a\in\vec{a}^i}(\sigma_n(a)-\eta_n(a))}{(1-\overline{\eta}^i_n)^{|\mathcal{I}^i|-1}}+\hat{\eta}^i_n(\vec{a}^i) & \text{otherwise.}\end{cases}$$

Note that

$$\begin{aligned}\sum_{\vec{a}^i\ni a}\hat{\sigma}^i_n(\vec{a}^i)&=\sum_{\vec{a}^i\ni a|\vec{a}^i\in\vec{A}^i_*}\frac{\prod_{a'\in\vec{a}^i}(\sigma_n(a')-\eta_n(a'))}{(1-\overline{\eta}^i_n)^{|\mathcal{I}^i|-1}}+\sum_{\vec{a}^i\ni a|\vec{a}^i\in\vec{A}^i_*}\hat{\eta}^i_n(\vec{a}^i)+\sum_{\vec{a}^i\ni a|\vec{a}^i\notin\vec{A}^i_*}\hat{\eta}^i_n(\vec{a}^i)\\ &=\sum_{\vec{a}^i\ni a|\vec{a}^i\in\vec{A}^i}\frac{\prod_{a\in\vec{a}^i}(\sigma_n(a')-\eta_n(a'))}{(1-\overline{\eta}^i_n)^{|\mathcal{I}^i|-1}}+\eta_n(a)\\ &=\frac{\sigma_n(a)-\eta_n(a)}{(1-\overline{\eta}^i_n)^{|\mathcal{I}^i|-1}}(1-\overline{\eta}^i_n)^{|\mathcal{I}^i|-1}+\eta_n(a)\\ &=\sigma_n(a)\ .\end{aligned}$$

Then, since each player plays once on the path of play in G^{sig}, it is clear that (1) $\hat{\sigma}^i_n(\vec{a}^i)\geq\hat{\eta}^i_n(\vec{a}^i)$ for all $\vec{a}^i\in\vec{A}^i$, and that (2) $\hat{\sigma}^i_n(\vec{a}^i)>\hat{\eta}^i_n(\vec{a}^i)$ only if $\vec{a}^i$ is optimal for i. Hence, $\hat{\sigma}_n\in\hat{\Sigma}^*_0(\hat{\eta}_n)$ (that is, is a Nash equilibrium of the normal-form of G^{sig} perturbed wih $\hat{\eta}_n$). Since $\hat{\sigma}_n$ generates the same outcome as σ_n, we have that ω is stable.

Part 2: Proof that if ω is stable in G^{sig}, then it is sequentially stable. In this part of the proof, we adapt the notation further to G^{sig}, as different arguments are made for the sender and the receiver. Now, $\vec{m}\equiv(\vec{m}_\theta\in M_\theta)_{\theta\in\Theta}$ and $\vec{r}\equiv(\vec{r}_m\in R_m)_{m\in M}$ denote normal-form pure strategies of the sender and the receiver, respectively. Let $\{I_\theta|\theta\in\Theta\}\subset\mathcal{I}$ and $\{I_m|m\in M\}\subset\mathcal{I}$ be the set of the information sets of the sender and the receiver, respectively. Consider a vanishing tremble $(\eta_n)_n$. Let $\eta_n(I):=\sum_{a\in A^I}\eta_n(a)$ be the sum of the trembles of information set I. We define $\overline{\eta}_n:=\max_\theta(\eta_n(I_\theta))$. Note that $\overline{\eta}_n\to 0$.

Fix a stable outcome ω. Abusing notation, we let $\omega(m|\theta)$ be the probability with which type θ sends message m under ω, and let $M^*_\theta:=\{m\in M_\theta|\omega(m|\theta)>0\}$. For each θ and message $m\in M_\theta$, we define

$$\eta'_n(m|\theta):=\begin{cases}\overline{\eta}_n^{1/|\Theta|-1}\,\eta_n(m|\theta) & \text{if } m\in M_\theta\setminus M^*_\theta,\\ \overline{\eta}_n^{1/|\Theta|-1}\,\omega(m|\theta)\big(\overline{\eta}_n-\sum_{m'\in M_\theta\setminus M^*_\theta}\eta_n(m')\big) & \text{if } m_\theta\in M^*_\theta.\end{cases}$$

Note that $\eta'_n(m|\theta)\searrow 0$ as $n\to\infty$ for all m and that $\eta'_n(I_\theta)=\overline{\eta}_n^{1/|\Theta|}$ for all $\theta\in\Theta$. Then, we define

$$\hat{\eta}^{\text{s}}_n(\vec{m}):=\prod_{\theta\in\Theta}\eta'_n(\vec{m}_\theta|\theta)\quad\text{and}\quad\hat{\eta}^{\text{r}}_n(\vec{r}):=\prod_{m\in M}\eta_n(\vec{r}_m|m)\ .$$

Since ω is stable, there is a sequence $(\hat{\sigma}_n)_n$ with the properties described above. Taking a subsequence if necessary (which, by Lemma A.1, is without loss for our argument), we assume that

$$\sigma(r|m) := \lim_{n\to\infty} \sum_{\vec{r}|\vec{r}_m=r} \hat{\sigma}_n(\vec{r}) \tag{A.4}$$

is well defined for all m and $r \in R_m$ (note that $\sum_{r\in R_m} \sigma(r|m) = 1$ for all m). Let $R_m^* := \{r \in R_m | \sigma(r|m) > 0\}$. For each n, define a behavior strategy profile $\sigma_n \in \Sigma$ as follows, for all $\theta \in \Theta$, $m \in M_\theta$, and $r \in R_m$:

$$\sigma_n(\theta) := \pi(\theta) , \quad \sigma_n(m|\theta) := \sum_{\vec{m}|\vec{m}_\theta=m} \hat{\sigma}_n^i(\vec{m}) , \quad \text{and}$$

$$\sigma_n(r|m) := \begin{cases} \eta_n(r|m) & \text{if } r \in R_m \backslash R_m^*, \\ K_n(m)\,\sigma(r|m) & \text{if } r \in R_m^*, \end{cases}$$

where $K_n(m)$ is chosen to be such that $\sum_{r\in R_m} \sigma_n(r|m) = 1$ for all $m \in M$. Note that

$$\begin{aligned} \sigma_n(m|\theta) \geq \sum_{\vec{m}|\vec{m}_\theta=m} \hat{\eta}_n^s(\vec{m}) &= \sum_{\vec{m}|\vec{m}_\theta=m} \prod_{\theta'\in\Theta_m} \eta_n'(\vec{m}_{\theta'}|\theta') = \Bigg(\prod_{\theta'\in\Theta_m\backslash\{\theta\}} \eta_n'(I_{\theta'}) \Bigg) \eta_n'(m|\theta) \\ &= \left(\overline{\eta}_n^{1/|\Theta|}\right)^{|\Theta|-1} \overline{\eta}_n^{-1/|\Theta|-1} \eta_n(m|\theta) = \eta_n(m|\theta) , \end{aligned}$$

where we used that $\eta_n'(I_\theta) = \overline{\eta}_n^{1/|\Theta|}$ by construction. Hence, since $\sigma_n(r|m) \geq \eta_n(r|m)$, we have that $\sigma_n \in \Sigma(\eta_n)$. Standard continuity arguments imply that, since $\hat{\sigma}_n$ is a Nash equilibrium of the normal-form game perturbed with $\hat{\eta}_n$, σ_n is asymptotically sequentially optimal for all types θ.[38] Then, ω is sequentially stable. □

Proof of Proposition 5.3

Proof. We prove the third part (NWBR_{CK}), as it is the strongest statement. The other parts can be proven similarly. Take then a sequentially stable outcome ω and an off-path message m. We let $(\check{\sigma}, \check{\mu})$ be a sequential equilibrium with outcome ω (which exists by Corollary 3.1). We let $\hat{\Theta} \subset \Theta_m$ be the set of types $\theta \in \Theta_m$ satisfying ($\text{NWBR}_{\textbf{CK}}$), that is, $\theta \in \hat{\Theta}$ if and only if, for all $\rho \in \text{BR}_m$ such that $u_\theta(m, \rho) = u_\theta(\omega)$, there is some $\theta' \in \Theta_m$ such that $u_{\theta'}(m, \rho) > u_{\theta'}(\omega)$. We assume $\hat{\Theta} \neq \Theta_m$.

Fix some $\mu_m \in \Delta(\Theta_m \backslash \hat{\Theta})$. Since ω is sequentially stable, Proposition 5.1 establishes that there are some $\alpha \in [0, 1]$, $\mu_m' \in \Delta(\Theta_m)$, and $\rho \in \text{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu_m')$, satisfying that $u_\theta(m, \rho) \leq$

[38] This is because the receiver's response to message m tends to $\sigma(\cdot|m) \in \Delta(R)$ defined in (A.4) under both sequences $(\sigma_n)_n$ and $(\hat{\sigma}_n)_n$, and since the belief of the receiver after each message m coincide under both sequences $(\sigma_n)_n$ and under $(\hat{\sigma}_n)_n$.

$u_\theta(\omega)$ for all $\theta \in \Theta_m$ and, if $\alpha \neq 1$, then $u_\theta(m,\rho) = u_\theta(\omega)$ for all $\theta \in \Theta_m$ with $\mu'_m(\theta) > 0$. If $\alpha = 1$, define $\hat{\alpha} := 1$, $\hat{\mu}'_m := \mu_m$, and $\hat{\rho} := \rho$. If $\alpha \neq 1$, then we argue that $\mu'_m(\theta) = 0$ for all $\theta \in \hat{\Theta}$. Indeed, assume for the sake of contradiction that $\mu'_m(\hat{\theta}) > 0$ for some $\hat{\theta} \in \hat{\Theta}$. In this case, by definition of $\hat{\Theta}$ and since $\rho \in \mathrm{BR}_m$, there is some type $\theta' \in \Theta_m$ such that $u_{\theta'}(m,\rho) > u_{\theta'}(\omega)$, but this contradicts that $u_{\theta''}(m,\rho) \leq u_{\theta''}(\omega)$ for all $\theta'' \in \Theta_m$. Define then, for the case $\alpha \neq 1$, $\hat{\alpha} := \alpha$, $\hat{\mu}'_m := \mu'_m$, and $\hat{\rho} := \rho$, and note that we have shown that $\hat{\mu}'_m \in \Delta(\Theta_m \backslash \hat{\Theta})$. It then follows that, for all $\mu_m \in \Delta(\Theta_m \backslash \hat{\Theta})$, there are $\hat{\alpha} \in [0,1]$, $\hat{\mu}'_m \in \Delta(\Theta_m \backslash \hat{\Theta})$, and $\hat{\rho} \in \mathrm{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\hat{\mu}'_m)$ such that the properties in Proposition 5.1 hold. Define $(\check{\sigma}', \check{\mu}')$ as

$$(\check{\sigma}'(\tilde{m}|\theta), \check{\sigma}'(r|\tilde{m}), \check{\mu}'_{\tilde{m}}(\theta)) := \begin{cases} (\check{\sigma}(\tilde{m}|\theta), \check{\sigma}(r|\tilde{m}), \check{\mu}_{\tilde{m}}(\theta)) & \text{if } \tilde{m} \neq m, \\ (\check{\sigma}(\tilde{m}|\theta), \hat{\rho}(r|\tilde{m}), \alpha\,\mu_m(\theta) + (1-\alpha)\,\hat{\mu}'_m(\theta)) & \text{if } \tilde{m} = m, \end{cases}$$

for all $\theta \in \Theta$, $\tilde{m} \in M_\theta$, and $r \in R_{\tilde{m}}$. It is then not difficult to see that $(\check{\sigma}', \check{\mu}')$ is a sequential equilibrium with outcome ω satisfying that $\check{\mu}'_m(\hat{\Theta}) = 0$. $\square$