ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 254

Sequentially Stable Outcomes

Francesc Dilmé

February 2024 (updated Version)

www.econtribute.de

UNIVERSITÄT BONN

Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Sequentially Stable Outcomes

Francesc Dilmé*

February 21, 2024

Abstract

This paper introduces and analyzes *sequentially stable outcomes* in extensive-form games. An outcome ω is sequentially stable if, for any $\varepsilon > 0$ and any small enough perturbation of the players' behavior, there is an ε-perturbation of the players' payoffs and a corresponding equilibrium with outcome close to ω. Sequentially stable outcomes exist for all finite games and are outcomes of sequential equilibria. They are closely related to stable sets of equilibria and 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 outcomes of stable sets of equilibria.

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 Lauermann, Massimo Marinacci, Georg Nöldeke, Lucas Pahl, Klaus Ritzberger, the associate editor (Marina Halac) and two anonymous referees, and seminar participants at Bocconi University, CERGE-EI in Prague, the Hebrew University of Jerusalem, the Humboldt University of Berlin, Royal Holloway University of London, Tel Aviv University, the University of Bonn, the University of Naples, the University of Texas at Austin, EEA-ESEM 2022 in Milan, and the Economic Theory Conference in Honor of George J. Mailath's 65th Birthday, for their useful comments. I also thank Marta Garriga-Massoni for her encouragement and support. This work was funded by a grant from the European Research Council (ERC 949465). Support from 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

Equilibrium refinements play a central role in the study of extensive-form games. Among these, the concept of a *sequential equilibrium*, introduced by Kreps and Wilson (1982), stands out because of its universality, simplicity, desirable properties (such as existence, belief consistency, and sequential rationality), and ease of use. Sequential equilibria have been widely used in analyzing games of incomplete information across a broad range of applications.

We start by providing a characterization of sequential outcomes (i.e., outcomes of sequential equilibria), presenting them as behavior that is robust to some tremble sequence: We show that an outcome ω is sequential if and only if there is some vanishing sequence of (behavioral) trembles (ξ_n) for which there exist a sequence $(\varepsilon_n)\to 0$ and a sequence of strategy profiles (σ_n) with outcomes converging to ω, such that each σ_n is a sequential ε_n-equilibrium of the game perturbed according to ξ_n.[1] Note that if one requires $\varepsilon_n = 0$ for all n instead of $\varepsilon_n \to 0$, our characterization of sequential outcomes coincides with the definition of outcomes of *perfect equilibria* in Selten (1975).

We then introduce and analyze a strengthening of sequentiality suggested by our characterization: We study behavior that is robust to *all* tremble sequences instead of just one. We say that an outcome ω is *sequentially stable* if, for any vanishing sequence of trembles (ξ_n), there exist a sequence $(\varepsilon_n)\to 0$ and a sequence of strategy profiles (σ_n) with outcomes converging to ω, such that each σ_n is a sequential ε_n-equilibrium of the game perturbed according to ξ_n.[2] This paper's main contribution is to propose sequential stability as a new equilibrium concept and show that it significantly strengthens sequentiality while keeping its ease of use. We relate sequentially stable outcomes to stable sets of equilibria (Kohlberg and Mertens, 1986) and also to the selection criteria for signaling games introduced in Cho and Kreps (1987).

We first establish that all extensive-form games have at least one sequentially stable outcome. To prove this, we perturb the payoffs of the agent-extensive form of the game. We then use that, for a generic payoff perturbation, there is a connected stable set of equilibria of the agent-extensive form of the game with the same outcome. We finally show that a limit of such outcomes for some vanishing sequence of generic payoff perturbations exists and is sequentially stable in the original game. This existence property is important not only because it ensures that sequential stability can

[1]A *behavioral tremble* assigns a minimal probability with which each action is played at each information set. We say that a strategy profile is a *sequential ε_n-equilibrium* of a game perturbed by a tremble if, whenever a player chooses to play an action with a probability higher than the tremble at a given information set, the payoff from playing such an action is at most ε_n lower than the payoff from the best response (see Definition 2.3).

[2]We show that this definition of sequentially stable outcome is equivalent to that in the abstract (see Proposition 4.2).

be used in all games, but also because it enables us to show that an outcome is sequentially stable by eliminating alternatives. Also, when an outcome is the unique limit equilibrium outcome along some tremble sequence, it is the unique sequentially stable outcome.

Next, we provide some properties satisfied by sequentially stable outcomes. First, 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 also sequentially stable in the game where a is removed. We use this result to show versions of both forward induction and iterated strict equilibrium dominance. We also establish 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 outcome can be replaced by that outcome without affecting the set of sequentially stable outcomes. Finally, we prove that the set of sequentially stable outcomes is invariant to how simultaneous moves are represented in the extensive form of the game. However, like sequential outcomes, sequentially stable outcomes may fail admissibility and may not be invariant to coalescing consecutive moves. Through several examples, 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), and we provide a full characterization of sequential stability without using trembles. 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 there is a unique joint outcome of a stable set of equilibria. We also show that the set of sequentially stable outcomes coincides with the set of outcomes of stable sets of equilibria in signaling games with generic payoffs.

Relationship to Kohlberg and Mertens (1986, KM): Sequential stability is closely related to the concept of stability, which from now we will refer to as "KM-stability," introduced by Kohlberg and Mertens (1986). Roughly speaking, a set of Nash equilibria is KM-stable if it is minimal with respect to the property that, for any vanishing sequence of normal-form trembles (assigning a positive probability to each contingent plan of each player), there is a sequence of Nash equilibria approaching the set. KM-stable sets of equilibria exist for all games and have desirable properties (they satisfy forward induction, iterated dominance, and invariance). Still, KM-stability is rarely applied as a selection criterion in practice. The main reason is that KM-stable sets of equilibria are difficult to characterize and use. In many extensive-form games, the number of sets of Nash equilibria that may be KM-stable a priori is large. While forward induction and iterated dominance allow some of them to be ruled out, it is typically infeasible to identify a KM-stable set by ruling out all alternatives. On the other hand, directly proving that a given set of equilibria is KM-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.[3]

Both sequential stability and KM-stability are based on the requirement of robustness to all small perturbations of the game. However, these concepts differ in two important ways. First, sequential stability requires ε-optimality along the sequence (for some $\varepsilon \to 0$) instead of exact optimality. This weakening permits the existence of an outcome-valued concept for all games, which is much easier to work with than a set-valued concept, but it also implies that sequential stability is not powerful in selecting equilibria in normal-form games, where all Nash outcomes are sequentially stable. Nevertheless, sequential stability offers significant selection power in extensive-form games; for example, it is stronger than the standard selection criteria in signaling games. The requirement of ε-optimality for some $\varepsilon \to 0$ also makes it easier to construct supporting equilibrium sequences, since, as with sequential equilibria, exact sequential optimality is only required in the limit. We are able to show that, if a game has a unique sequentially stable outcome, such an outcome is the limit of a sequence of Nash outcomes (i.e., with $\varepsilon_n = 0$ for all n) along any sequence of vanishing trembles. The second main difference between sequential stability and KM-stability is that the former applies to the extensive form of the game, instead of the reduced normal form. This permits us to use simpler, more intuitive arguments. It also lets us apply methods such as NWBR or strict domination action by action, instead of considering full contingent plans, and it lets us simplify the analysis by replacing subgames with their sequentially stable outcomes.

Contribution to the literature: Since the definition of Nash equilibria (Nash, 1951), many equilibrium concepts have been developed to select 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), KM-stable sets (Kohlberg and Mertens, 1986), and perfect Bayesian equilibria (Fudenberg and Tirole, 1991b).[4]

[3] Okada (1981) defined *strictly perfect equilibria* as those that are robust against all perturbations. However, such equilibria do not exist in many games of interest, even when payoffs are generic. Kohlberg and Mertens (1986) show that "there exists a [KM-]stable set which is contained in a single connected component of the set of Nash equilibria" (p. 1027) and that, generically in payoffs, "all equilibria in the same connected component give rise to identical outcomes" (p. 1020). Nevertheless, joint outcomes of KM-stable sets contained in connected sets of equilibria are difficult to compute, and cannot be used as a universal equilibrium concept because many games of interest do not have generic payoffs due to quasilinear preferences, payoff-irrelevant signals, assumed functional forms, or constant discount factors.

[4] Sometimes, when a given equilibrium concept is not powerful enough as a selection criterion, additional ad-hoc require-

In signaling games, selection criteria such as the Intuitive Criterion, D1, and D2 of Cho and Kreps (1987) or the divinity criterion of Banks and Sobel (1987) may be used. This variety of concepts has made it increasingly difficult to compare equilibrium predictions across different games.[5]

We contribute to the literature by providing a new equilibrium concept that is 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 and always exist. Second, sequential stability is stronger than numerous previous equilibrium concepts, including subgame-perfect, sequential, and perfect Bayesian equilibria, and it passes commonly used selection criteria in signaling games. It can therefore be used directly with previous work involving unique solutions: for example, any game having a unique sequential outcome, or a unique outcome passing D1, automatically has a unique sequentially stable outcome. Third, the properties of sequentially stable outcomes—such as NWBR and forward induction, which are defined through natural conditions on the optimality of actions in each information set instead of on the global optimality of full contingent plans—make them easier to compute. We provide examples illustrating this throughout the paper. In a companion paper, Dilmé (2023b), 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. In Section 2 we establish our notation for extensive-form games, define vanishing trembles and sequential ε-equilibria, and provide a new characterization of sequential equilibria. In Section 3 we define sequentially stable outcomes, relate them to KM-stable sets of equilibria, and prove that all games have a sequentially stable outcome. In Section 4 we obtain properties of sequentially stable outcomes and describe techniques for finding them. In Section 5 we characterize sequential stability in signaling games and show that sequentially stable outcomes pass common selection criteria. Finally, Section 6 concludes. The appendix contains the proofs of the results.

---

ments are imposed, 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, which is studied in Fudenberg et al. (1988). Recently, Takahashi and Tercieux (2020) have shown the existence of outcomes robust to payoff uncertainty for generic payoffs.

2 Sequential ε-equilibria and sequential outcomes

2.1 Extensive-form games

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

A (finite) *extensive-form game* $G := \langle A, H, \mathcal{I}, N, \iota, \pi, u\rangle$ has the following components: (1) A finite set of *actions* A. (2) A finite set of *histories* H containing finite sequences of actions such 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 by Z. (3) An *information partition* $\mathcal{I}$ of the set of non-terminal histories such that there is a partition $\{A^I | I \in \mathcal{I}\}$ of A with the property that, for each $I \in \mathcal{I}$ and $h \in H$, we have $(h, a) \in H$ for some $a \in A^I$ if and only if $h \in I$.[6] (4) A finite set of *players* $N \not\ni 0$. (5) A *player assignment* $\iota : \mathcal{I} \to N \cup \{0\}$ assigning each information set to a player or nature such that there is perfect recall.[7] (6) 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 $I \in \iota^{-1}(\{0\})$. (7) For each player $i \in N$, a *(von Neumann–Morgenstern) payoff function* $u_i : Z \to \mathbb{R}$. For convenience, we set $u_0(z) = 0$ for all $z \in Z$.

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. An *outcome* ω (of G) is a probability distribution over terminal histories. We use $\Omega := \Delta(Z)$ to denote the set of outcomes. Each strategy profile $\sigma \in \Sigma$ generates a unique outcome ω^σ, where each terminal history $(a_j)_{j=1}^J \in Z$ is assigned probability $\omega^\sigma((a_j)_{j=1}^J) := \prod_{j=1}^J \sigma(a_j) \in [0, 1]$.

2.2 Trembles and vanishing trembles

Next we define trembles and vanishing trembles. 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, that is, satisfying $a \in A^{I^a}$.

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

As is common, we interpret $\xi(a) \in (0, 1]$ as the smallest probability with which player $\iota(I^a)$

[6] Note that we assume, without loss of generality, that each action is available at a unique information set (otherwise, one can rename actions).

[7] Perfect recall requires that for all $I, I' \in \mathcal{I}$ with $\iota(I) = \iota(I')$ and all $h, \hat{h} \in I$, if $(h', a) \preceq h$ for some $h' \in I'$ and $a \in A$, then $(\hat{h}', a) \preceq \hat{h}$ for some $\hat{h}' \in I'$, where $(h', a) \preceq h$ indicates that (h', a) preceeds or is equal to h.

can decide to select action $a \in A$. A tremble thus represents the probability with which players make mistakes. We denote by $\Sigma(\xi)$ the set of strategy profiles $\sigma \in \Sigma$ such that, for all $a \in A$, $\sigma(a) \geq \xi(a)$. For each tremble, $G(\xi)$ denotes the *perturbed game* defined by G together with the set of strategy profiles $\Sigma(\xi)$. As we are interested in small trembles, we will often work with vanishing trembles.

Definition 2.2. A *vanishing tremble* is a sequence of trembles (ξ_n) such that $\xi_n(a) \to 0$ for all $a \in A$.

A vanishing tremble (ξ_n) generates a sequence of perturbed games $(G(\xi_n))$. Such a sequence approaches G (with the set of strategy profiles Σ), in the sense that the sets of strategy profiles $\Sigma(\xi_n)$ approach Σ (under the Hausdorff distance) as $n \to \infty$.[8]

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

Fix a tremble ξ. Note that all information sets are reached with positive probability under any strategy profile $\sigma \in \Sigma(\xi)$. Then, for each action $a \in A$, the expected payoff of player $\iota(I^a)$ from playing a conditional on I^a being reached, is uniquely defined. This payoff is

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

where $Z^a \subset Z$ 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.

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

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

[8] 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 and sup-distance.

2.4 A characterization of sequential outcomes

Kreps and Wilson (1982) defined *belief system* μ as a map assiging a probability $\mu(h)\in[0,1]$ to each non-terminal history $h\in H\backslash Z$ satisfying that $\sum_{h\in I}\mu(h)=1$ for all $I\in\mathcal{I}$. They also defined a *sequential equilibrium* as a pair consisting of a belief system μ and a strategy profile σ that is consistent and sequentially rational. Part of their motivation was to make the new equilibrium concept similarly powerful but simpler to use than the concept of *(trembling-hand) perfect equilibria* (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). The following characterization of sequential outcomes brings the two concepts closer together.

Proposition 2.1. *An outcome ω is sequential if and only if there exist a vanishing tremble (ξ_n), a sequence $(\varepsilon_n)\to 0$, and a sequence $(\sigma_n\in\Sigma^*_{\varepsilon_n}(\xi_n))$ such that (ω^{σ_n}) converges to ω.*

This proposition makes it evident that sequential equilibria are a weakening of perfect equilibria (as shown by Kreps and Wilson, 1982), requiring only ε_n-optimality along the sequence for some $(\varepsilon_n)\to 0$ instead of exact optimality (i.e., $\varepsilon_n=0$). 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 a characterization of sequential equilibria in terms of conditional ε-equilibria.[9]

Example 2.1. Figure 1 shows the *beer–quiche game* of Cho and Kreps (1987). In this game, nature chooses whether player 1's type is strong or weak; then player 1 chooses either beer or quiche, and player 2, observing player 1's choice but not her type, chooses either to fight or to run. We now use an explicit vanishing tremble to show that the *quiche outcome* ω_{q}—in which player 1 chooses $\mathrm{q_s}$ and $\mathrm{q_w}$, and then player 2 chooses $\mathrm{r_q}$—is sequential. Let (ξ_n) satisfy $\xi_n(\mathrm{b_s})=\xi_n(\mathrm{q_s})=(n+1)^{-2}$ and $\xi_n(a)=(n+1)^{-1}$ for all $a\neq\mathrm{b_s},\mathrm{q_s}$. Note that the weak type trembles with an asymptotically infinitely higher likelihood than the strong type. Consider the sequence of strategy profiles (σ_n) pinned down by $\sigma_n(a):=\xi_n(a)$ for all $a\in\{\mathrm{b_s},\mathrm{b_w},\mathrm{r_b},\mathrm{f_q}\}$. It is then easy to see that $\sigma_n\in\Sigma^*_0(\xi_n)$ for all large enough n and that σ_n tends to ω_{q}. Hence, by Proposition 2.1, the quiche outcome is sequential. (Note that the beer–quiche game is simple enough to allow us to construct a sequence

[9] Myerson and Reny (2020) show that ω is sequential if and only if, for some vanishing tremble, it is the limit of a corresponding sequence of outcomes of conditional ε-equilibria for some $\varepsilon\to 0$, where they define σ to be a *conditional ε-equilibrium* if $\sum_{a\in A^I}\sigma(a)u(a|\sigma)\geq\max_{\hat\sigma_I\in\Delta(A^I)}\sum_{a\in A^I}\hat\sigma_I(a)u(a|\sigma)-\varepsilon$ for all $I\in\mathcal{I}$. It is not difficult to verify that, in fact, ω is sequential if and only if, for *any* vanishing tremble, it is the limit of a corresponding sequence of outcomes of conditional ε-equilibria for some $\varepsilon\to 0$. As we shall see, requiring instead that ω is the limit of outcomes of sequential ε-equilibria for some $\varepsilon\to 0$ along all vanishing trembles will significantly refine the set of sequential outcomes.

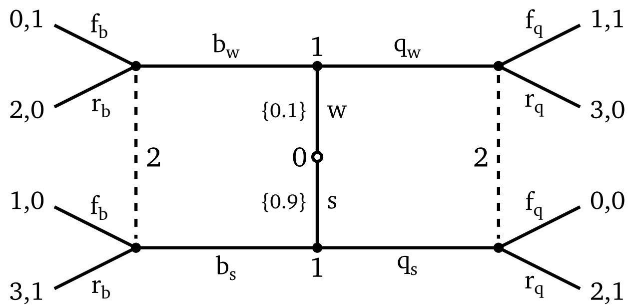

Figure 1

that is exactly optimal for all n, hence ω_q is also perfect.) Intuitively, along the sequence of strategy profiles, the probability that player 2 assigns to the strong type after observing beer tends to 0, while the probability she assigns to the strong type after observing quiche tends to 0.9. This makes r_q and f_b asymptotically optimal for player 2, and hence q_s and q_w are asymptotically optimal for player 1.

3 Sequentially stable outcomes

In this section, we introduce the concept of sequentially stable outcomes and prove that they exist in any game. We also discuss their relationship to KM-stable sets of equilibria (Kohlberg and Mertens, 1986; see van Damme, 1991, for a textbook treatment).

3.1 Definition of sequentially stable outcomes

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 (ξ_n), there are two sequences $(\varepsilon_n) \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n))$ such that (ω^{σ_n}) converges to ω.

In words, an outcome is sequentially stable if, for any vanishing tremble, it can be approximated (under the sup-distance; see Footnote 8) by a sequence of sequential epsilon-outcomes of the corresponding perturbed games, for some vanishing sequence of epsilons. Our definition of sequentially stable outcomes is analogous to the characterization of sequential outcomes in Proposition 2.1 but requires robustness for all vanishing trembles instead of for one 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 3.1 (continuation of Example 2.1). In Example 2.1, we showed that the quiche outcome ω_{q} is sequential by showing it is robust to a particular vanishing tremble. We now prove that ω_{q} is *not* sequentially stable by showing it is *not* robust to another vanishing tremble. Consider the vanishing tremble (ξ_n) where $\xi_n(\mathrm{b_w})=\xi_n(\mathrm{q_w})=(n+1)^{-2}$ and $\xi_n(a)=(n+1)^{-1}$ for all $a\neq \mathrm{b_w},\mathrm{q_w}$. Note that now the strong type trembles with an asymptotically infinitely higher likelihood than the weak type. Assume for the sake of contradiction that there are two sequences $(\varepsilon_n)\to 0$ and $(\sigma_n\in\Sigma^*_{\varepsilon_n}(\xi_n))$ with $\omega^{\sigma_n}\to\omega_{\mathrm{q}}$. If there were a strictly increasing sequence (k_n) with $\sigma_{k_n}(\mathrm{b_w})=\xi_{k_n}(\mathrm{b_w})$, then we would have $\sigma_{k_n}(\mathrm{f_b})=\xi_{k_n}(\mathrm{f_b})$ for n large enough: because beer would become an increasingly strong signal that player 1 is strong, player 2 would respond to it by running. Thus, the strong type's payoff from choosing $\mathrm{b_s}$ would tend to 3 along this subsequence, which is strictly higher than her payoff from choosing $\mathrm{q_s}$, contradicting the assumption that choosing $\mathrm{q_s}$ is asymptotically optimal (i.e., there cannot be some $(\varepsilon_n)\to 0$ such that $u(\mathrm{q_s}|\sigma_n)\geq u(\mathrm{b_s}|\sigma_n)-\varepsilon_n$ for all n). We conclude that if n is large enough, $\sigma_n(\mathrm{b_w})>\xi_n(\mathrm{b_w})$, and so $\mathrm{b_w}$ must asymptotically optimal. However, since $\sigma_n(\mathrm{r_q})\to 1$ as $n\to\infty$ in the quiche outcome, the weak type's payoff from choosing $\mathrm{q_w}$ (which converges to 3) is larger than her payoff from choosing $\mathrm{b_w}$ (which is at most 2), contradicting the assumption that choosing $\mathrm{b_w}$ is asymptotically optimal. Intuitively, choosing quiche is asymptotically optimal for the strong type only if player 2 fights after beer with a high enough probability. This means player 2's asymptotic posterior about player 1 being strong after beer must be lower than 0.5; hence, choosing beer must be asymptotically optimal for the weak type. But the weak type obtains 3 by choosing quiche, which is more than her payoff from choosing beer.

Example 3.2 (continuation of Example 3.1). We now prove that the *beer outcome* ω_{b}—in which player 1 chooses $\mathrm{b_s}$ and $\mathrm{b_w}$, and then player 2 chooses $\mathrm{r_b}$— is sequentially stable. Fix an arbitrary vanishing tremble (ξ_n). For each n with $\xi_n(\mathrm{q_w})\geq 9\,\xi_n(\mathrm{q_s})$ define $\sigma_n(a):=\xi_n(a)$ for all $a\in\{\mathrm{q_s},\mathrm{q_w},\mathrm{r_q},\mathrm{f_b}\}$, which pins down the value of $\sigma_n(a)$ for all a. Note that, under such σ_n, if n is large enough, then both types of player 1 strictly lose from choosing quiche, and player 2 assigns a probability greater than or equal to 0.5 to $(\mathrm{s},\mathrm{q_s})$ after quiche. For each n with $\xi_n(\mathrm{q_w})<9\,\xi_n(\mathrm{q_s})$, define $\sigma_n(\mathrm{q_w}):=9\,\xi_n(\mathrm{q_s})$, $\sigma_n(\mathrm{r_q}):=0.5-\xi_n(\mathrm{f_b})$, and $\sigma_n(a):=\xi_n(a)$ for $a\in\{\mathrm{q_w},\mathrm{f_b}\}$, which again pins down the value of $\sigma_n(a)$ for all a. Now, under σ_n and for n 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 of 0.5 to $(\mathrm{s},\mathrm{q_s})$ after quiche. It is easy to see that $\sigma_n\in\Sigma^*_0(\xi_n)$ for all

n large enough and that $\omega^{\sigma_n} \to \omega_{\mathrm{b}}$; hence, ω_{b} is sequentially stable.[10]

A characterization of sequential stability

To give further intuition for sequential stability, we now characterize sequentially stable outcomes without using vanishing trembles: An outcome ω is sequentially stable if, for all $\varepsilon > 0$, any slightly perturbed version of G has a sequential 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 nearby almost-optimal outcome.

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 $\|\xi\| < \delta$, then $G(\xi)$ has a sequential ε-equilibrium with outcome ε'-close to ω.*

Extensive-form stable outcomes

It will often be useful to consider the following natural strengthening of sequential stability: We say that ω is *extensive-form stable* if it satisfies Definition 3.1 with the additional requirement that $\varepsilon_n = 0$ for all n. That is, if ω is extensive-form stable, then, for any vanishing tremble, there is a sequence of Nash outcomes of the corresponding perturbed games converging to ω. It is clear that extensive-form stability is stronger than sequential stability, as it requires $\varepsilon_n = 0$ instead of $\varepsilon_n \to 0$. We record this observation in the following proposition.

Proposition 3.2. *Every extensive-form stable outcome is sequentially stable.*

The following result establishes that, while sequential stability is weaker than extensive-form stability, the two concepts coincide when there is a unique sequentially stable outcome.

Proposition 3.3. *If there is a unique sequentially stable outcome, it is the unique extensive-form stable outcome.*

The intuition for Proposition 3.3 is as follows. If there is a unique sequentially stable outcome ω, then, by Proposition 3.2, either ω is the unique extensive-form stable outcome (in which case the result holds), or the game has no extensive-form stable outcome. If the latter, then there is a vanishing tremble with no corresponding sequence of Nash equilibria converging to ω. Combining

[10] It is also easy to see that in the beer–quiche game, the sets of limits of sequences of sequential ε_n-equilibria for different vanishing trembles may have empty intersection. This fact and Proposition 3.4 motivates applying sequential stability (i.e., robustness to small trembles) to outcomes instead of to strategy profiles to guarantee existence in all games.

this vanishing tremble with a sequence of perturbations of payoffs, we can construct a sequence of sequentially stable outcomes of nearby games converging to an outcome ω' different from ω. However, we show that the correspondence that maps payoffs to the set of sequentially stable outcomes is upper hemicontinuous, and so ω' must be sequentially stable. This contradicts the assumption that ω is the unique sequentially stable outcome.

3.2 Existence of sequentially stable outcomes

We now prove that sequentially stable outcomes always exist.

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

The proof of Proposition 3.4 is divided into two steps. We first argue that the agent-extensive form of G, denoted by $\hat{G}$, has an outcome that is the limit of extensive-form stable outcomes along a sequence of generically perturbed payoffs. To prove this result, we first argue that, if the payoffs of $\hat{G}$ are generically perturbed, it has an extensive-form stable outcome.[11] We then take a generic sequence of payoff functions $(\hat{u}_k : Z \to \mathbb{R}^{\hat{N}})$ converging to $\hat{u}$ and, for each k, an extensive-form stable outcome ω_k of $\hat{G}$ with payoffs $\hat{u}_k$. Taking a subsequence if necessary, we may assume that (ω_k) converges to some $\omega \in \Omega$. The second step of the proof shows that ω is sequentially stable in G. In this step, we take a vanishing tremble (ξ_n) and fix, for each k, a sequence $(\sigma_{k,n})$ with outcomes converging to ω_k and with $\sigma_{k,n} \in \Sigma_0^*(\xi_n, \hat{u}_k)$ for all n, where $\Sigma_\varepsilon^*(\xi_n, \hat{u}_k)$ indicates the set of sequential ε-equilibria of $\hat{G}$ perturbed according to ξ_n with payoff $\hat{u}_k$. (Such a sequence exists because ω_k is extensive-form stable.) We use a standard diagonal argument to prove that there exist an increasing sequence (n_k) and a sequence $(\varepsilon_k) \to 0$ such that $\sigma_{k,n_k} \in \Sigma_{\varepsilon_k}^*(\xi_{n_k}, \hat{u})$ for all k and $\omega^{\sigma_{k,n_k}}$ converges to ω as $k \to \infty$. Since the argument holds for any vanishing tremble, we argue that ω is sequentially stable in G.

The existence of a sequentially stable outcome in all games contrasts with the fact that KM-stable sets of equilibria with a common outcome 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 3.2 is not true in

[11]This follows from the results in Kohlberg and Mertens (1986) provided in Footnote 3. Hence, using that the sets of mixed and behavior trembles of $\hat{G}$ coincide, we have that, for a generic payoff perturbation of $\hat{G}$, the joint outcome of a KM-stable set of equilibria contained in a connected set is extensive-form stable. Note that this observation does not imply that extensive-form stable outcomes exist in games with generic payoffs; instead, implies that they exist in games with generic payoffs that coincide with their agent-extensive form.

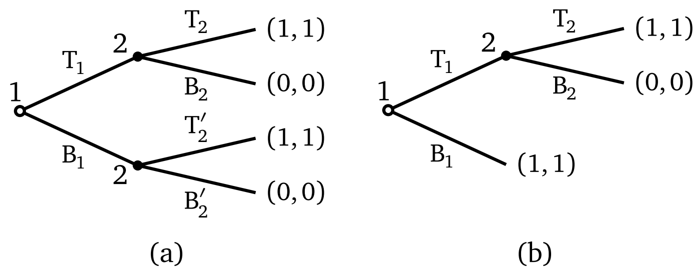

Figure 2

general: while all games have sequentially stable outcomes, some may not be common outcomes KM-stable sets of equilibria.

Example 3.3 (continuation of Example 3.2). Cho and Kreps (1987) show that the beer–quiche game has two sequential outcomes (the beer and the quiche outcomes described in Exercises 2.1 and 3.2). Because, by Example 3.1, the quiche outcome is not sequential, Proposition 3.4 implies that the beer outcome is the unique sequentially stable outcome. We now explicitly show that the beer outcome is the unique sequentially stable outcome by showing it is the unique limit equilibrium outcome along a particular vanishing tremble. We consider the vanishing tremble (ξ_n) given by $\xi_n(\mathrm{b_w})=\xi_n(\mathrm{q_w})=(n+1)^{-2}$ and $\xi_n(a)=(n+1)^{-1}$ for all $a\neq \mathrm{b_w}, \mathrm{q_w}$ (which is also considered in Example 3.1). Note that the strong type trembles asymptotically infinitely more than the weak type. Let $(\varepsilon_n)\to 0$ and $(\sigma_n\in \Sigma^*_{\varepsilon_n}(\xi_n))$ such that (ω^{σ_n}) converges to some outcome ω. If $\sigma_{k_n}(\mathrm{b_w})=\xi_{k_n}(\mathrm{b_w})$ along a sequence (k_n), then player 2 assigns a vanishing probability to $(\mathrm{w}, \mathrm{b_w})$ after beer along this sequence, hence it must be that $\sigma_{k_n}(\mathrm{r_b})\to 1$. It then follows that $\sigma_{k_n}(\mathrm{b_s})\to 1$, which necessarily implies that $\sigma_{k_n}(\mathrm{b_w})\to 1$, contradicting that $\sigma_{k_n}(\mathrm{b_w})=\xi_{k_n}(\mathrm{b_w})$ for all n. It must then be that $\sigma_n(\mathrm{b_w})>\xi_n(\mathrm{b_w})$ for all n large enough. Then, because the weak type asymptotically weakly prefers beer to quiche, the strong type strictly prefers beer to quiche for n large enough. Again, this implies $\sigma_n(\mathrm{b_s})\to 1$ and $\sigma_n(\mathrm{b_w})\to 1$, so ω is the beer outcome. Because the beer outcome is the unique limit equilibrium outcome for (ξ_n), it is the unique sequentially stable outcome of the beer–quiche game. By Proposition 3.3, it is also the unique extensive-form stable outcome.

Example 3.4. The proof of Proposition 3.4 illustrates how the requirement of ε_n-optimality (instead of exact optimality) in the definition of sequential stability permits us to simultaneously perturb a game's strategies and payoffs to show that the limit of a sequence of extensive-form stable outcomes of nearby games is a sequentially stable outcome. Such an argument would not apply to extensive-form stable outcomes, which may not exist in games with non-generic payoffs. To see this, consider the game in Figure 2(a), which corresponds to Figure 6.4.1 in van Damme (1991). Fix an outcome ω assigning a positive probability to action T_1. Consider a vanishing tremble (ξ_n) such that $\xi_n(\mathrm{B}_2)>$

$\xi_n(B_2')$, that is, such that player 2 trembles more toward B_2 than toward B_2'. Because it is optimal for player 2 to choose T_2 and T_2' in any Nash equilibrium for any n, player 1 prefers choosing B_1 (which gives her a payoff of $1-\xi_n(B_2')$) to choosing T_1 (which gives her $1-\xi_n(B_2)$). Hence, $\sigma_n(B_1)\to 1$ in any sequence of Nash equilibria along games perturbed according to (ξ_n), and so ω is *not* extensive-form stable. A symmetric argument implies that an outcome assigning positive probability to B_1 is not extensive-form stable, so there is no extensive-form stable outcome. Note that, nevertheless, player 1's payoffs from playing T_1 and B_1 are approximately the same, because player 2 chooses T_2 and T_2' with asymptotic probability 1. Thus, it is easy to see that any outcome in which player 2 responds with T_2 or T_2' to player 1's on-path actions is sequentially stable. It is also easy to see that an outcome in which player 1 fully mixes is *not* the limit of extensive-form stable outcomes of nearby games with slightly perturbed payoffs.

3.3 Relationship to KM-stable sets of equilibria

The definition of a sequentially stable outcome differs from the definition of a KM-stable set of equilibria in Kohlberg and Mertens (1986) in two important ways.[12] The first difference is that sequential stability requires only 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 mixed strategies (i.e., a player's tremble assings a postive probability to all her full contingent plans), while we consider independent trembles to the actions in each information set. We see these two departures as necessary to produce a single-valued equilibrium concept that exists in all games, has high selection power, and possesses desirable properties that permit one to consider incentives related to actions instead of incentives related to full contingent plans. 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 increasing tractability and ensuring the existence of sequentially stable outcomes in all games. It is analogous to the relaxation of the exact optimality of perfect equilibria to the ε_n-optimality of sequential equilibria established in Proposition 2.1.[13] While this type of relaxation preserves significant selection

[12]Roughly speaking, a set of equilibria S is *KM-stable* if it is minimal with respect to the property of being closed and such that, for any vanishing sequence of normal-form trembles (each assigning minimal probability to each contingent plan of each player), there is a corresponding sequence of Nash equilibria of the perturbed games approaching S.

[13]While the weakening provided by sequential equilibria was not needed for existence (existence of perfect equilibria had been established in Selten, 1975), it made verifying properties (such as sequential optimality) much easier, as they could be verified directly "in the limit". The same applies to sequentially stable outcomes; for example, Dilmé (2023b)

power in extensive-form games, the same approach does not work for 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 such that, for any vanishing tremble (ξ_n), there exist two sequences $(\varepsilon_n)\to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n))\to\sigma$. The implication is that all Nash outcomes of a normal-form game are sequentially stable (and hence sequential outcomes).[14]

Given our requirement of approximate sequential optimality along the sequence, it is natural to focus on behavioral vanishing trembles. As Corollary 3.1 establishes, the requirement of robustness 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, giving an additional sense of robustness. In contrast, sequential rationality is not guaranteed in KM-stable sets of equilibria. For instance, Kohlberg and Mertens (1986) show that KM-stable sets do not necessarily satisfy sequential rationality by providing an example of a game with a KM-stable set of equilibria having a common outcome that is not the outcome of the unique sequential equilibrium (see their Figure 11). Furthermore, van Damme (1991) exhibits a game with a KM-stable set of equilibria having a common outcome which is not the outcome of the unique subgame-perfect Nash equilibrium (see his Example 10.3.4).[15] By Corollary 3.1, the unique sequentially stable outcome of the games in these examples is the outcome of their unique sequential equilibrium.

4 Properties of sequentially stable outcomes

In this section, we provide some properties that sequentially stable outcomes satisfy, and we compare them with the properties of KM-stable sets of equilibria. We also provide some examples showing how these properties are used.

---

characterizes sequential stability directly at the limit.

[14] 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. By contrast, Takahashi and Tercieux (2020) show that requiring outcomes to be robust to payoff perturbations has significant selection power in extensive-form games. However, robust outcomes do not exist for all games.

[15] Mertens (1989) provides a definition of KM-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 KM-stable set contains a sequential equilibrium). However, Mertens's definition is remarkably involved and difficult to use in practice.

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 implies both 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 in which a is removed (as are all histories following it).*[16]

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 by assigning to a a sequence of probabilities that vanish much faster than the probabilities assigned by the vanishing tremble to any other action. Take a corresponding sequence of sequential ε_n-equilibria with outcomes converging to ω (which exists because ω is sequentially stable). The proof of Proposition 4.1 shows that the restrictions of sequential ε_n-equilibria to $\hat{G}$ generate a sequence of sequential ε_n-equilibria, for some sequence $(\hat{\varepsilon}_n) \to 0$. Intuitively, since a is not sequentially optimal under any sequential equilibrium with outcome ω, each of the sequential ε_n-equilibria for n high enough assigns to a the same very low probability as it has in the vanishing tremble. As a result, any history containing a has a vanishing likelihood relative to any history not containing a, ensuring the asymptotic sequential rationality of the restrictions of the sequential ε_n-equilibria to $\hat{G}$.

Our definition of NWBR implies the following versions of forward induction and iterated strict equilibrium dominance.

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 removed, ω remains sequentially stable.

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

[16] An action a 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 maximum continuation payoff that player $\iota(I^a)$ can obtain by playing some action in A^{I^a}. Note that ω is an outcome of any game that results from eliminating an action that is off path under ω.

Forward induction and iterated strict equilibrium dominance are intuitive and often easier to use than NWBR.[17] Forward induction arguments can be used to rule out candidates for sequentially stable outcomes by proving they are not sequentially stable in a simpler game. 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 a candidate outcome.

Example 4.1. Before, we argued that KM-stability does not imply sequential stability, as there are games with KM-stable sets of equilibria that do not contain sequential equilibria. Figure 3(a), which coincides with Figure 2 in Kohlberg and Mertens (1986), provides an example showing that sequential stability does not imply KM-stability. Assume $x \in (0,1)$. In this game, the outcome assigning probability one to T_1 is not the outcome of all equilibria in a KM-stable set.[18] Still, such an outcome is sequentially stable: For any vanishing tremble (ξ_n) and any n, one can define $\sigma_n(a) := \xi_n(a)$ for all $a \in \{\mathrm{B}_1', \mathrm{M}_1, \mathrm{T}_2\}$ (which pins down the full strategy profile), and this supports the outcome. Intuitively, player 2 interprets the choice of B_1' as a mistake; hence, she believes that player 1 will play B_1 because this is optimal given the prescribed strategy profile.

Example 4.2 (continuation of Example 3.3). Kohlberg and Mertens (1986) argue that the beer outcome ω_{b} is the unique outcome of a KM-stable set of equilibria of the beer–quiche game (see Figure 1). They do so by claiming that the set of Nash equilibria has two connected components, ruling out the KM-stability of one component using forward induction, and finally claiming that the other component must contain a KM-stable set. A complete argument would require proving that there are two connected components of Nash equilibria and that there are no KM-stable sets containing equilibria in both connected components (note that, as defined by Kohlberg and Mertens, a KM-stable set need not be connected). To characterize the possible KM-stable sets, one would then need to impose minimality.

[17]Note that the conclusion of Corollary 4.1 also holds if the left-hand side of (4.1) is replaced by $\max_{z \in Z^a} u_{\iota(I)}(z)$ (that is, if 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 one need not 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 $\max_{z \in Z^a} u_{\iota(I)}(z) < \min_{z \in Z^{a'}} u_{\iota(I)}(z)$, then ω remains sequentially stable if a is eliminated (recall that Z^a is the set of terminal histories that contain a).

[18]To see this, assume for the sake of contradiction that there is a KM-stable set of equilibria of the reduced normal-form game (where players 1 and 2 have action sets $\{\mathrm{T}_1, \mathrm{B}_1'\mathrm{M}_1, \mathrm{B}_1'\mathrm{B}_1\}$ and $\{\mathrm{T}_2, \mathrm{B}_2\}$, respectively) with outcome assigning probability one to T_1. Because $\mathrm{B}_1'\mathrm{B}_1$ is an inferior response in all equilibria with outcome assigning probability one to T_1, the KM-stable set contains a KM-stable set of the game obtained by deleting $\mathrm{B}_1'\mathrm{B}_1$ (by Proposition 6 in Kohlberg and Mertens, 1986). Action B_2 can be eliminated from the resulting game using a similar argument. Nevertheless, it is easy to see that there is no KM-stable set with outcome assigning probability one to T_1 in the game without $\mathrm{B}_1'\mathrm{B}_1$ and B_2.

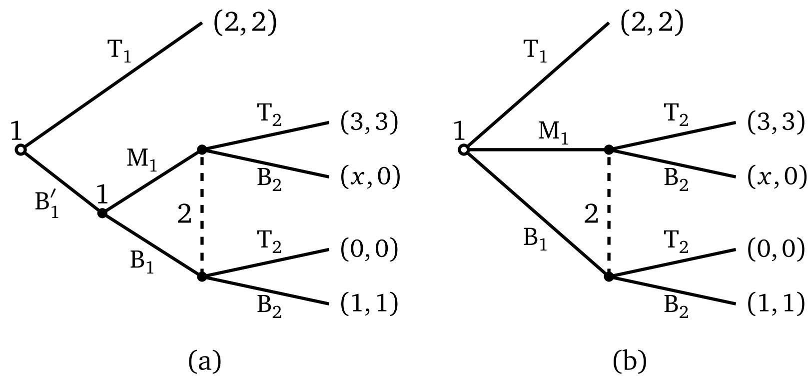

Figure 3

We now provide a similar but simpler argument using forward induction and equilibrium dominance, which is both straightforward and complete (see Examples 3.1–3.3 for analogous results using vanishing trembles). We start by showing that the quiche outcome ω_q is *not* sequentially stable. If it were, then 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 b_w is lower than her payoff under ω_q. In the game without action b_w, action f_b is strictly (equilibrium) dominated, so it can also be eliminated. In the resulting game, the strong type prefers playing b_s (which can only be followed by r_b) to playing q_s, so there is no sequential equilibrium with the quiche outcome, contradicting the assumption of its sequential stability.

Next we show that the beer outcome is the only remaining candidate for a sequential outcome. To see this, take a sequential equilibrium (σ, μ). It cannot be that $\sigma(b_w) \in (0,1)$, since then $\sigma(r_b) = \sigma(r_q) + 1/2$ and the strong type strictly prefers b_s to q_s, leading to $\sigma(r_b) = 1$ and $\sigma(r_q) = 0$, so that the weak type strictly prefers b_w to q_w. Similarly, it cannot be that $\sigma(b_s) \in (0,1)$, since then $\sigma(r_b) + 1/2 = \sigma(r_q)$ and the weak type strictly prefers q_w to b_w, leading to $\sigma(r_b) = 1$ and so $\sigma(r_q) = 3/2$. It must then be that $\sigma(b_w), \sigma(b_s) \in \{0,1\}$. If $\sigma(b_w) \neq \sigma(b_w)$ then the weak type is fought on path while the strong type is not, so the weak type has an incentive to deviate. Hence, the beer outcome is the only candidate for a sequentially stable outcome. By Proposition 3.4, it follows that the beer outcome is the unique sequentially stable outcome.

Remark 4.1. Note that 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 KM-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 any sequential equilibria with that outcome. Our definition of NWBR is thus applied to a single-valued object (outcomes instead of sets of equilibria) and requires simpler conditions (on actions instead of full contingent plans).[19] Similarly, our version of forward induction permits us to eliminate actions that are available on path but are strictly dominated by not deviating (in the sense of (4.1)). We see it as natural to require the actions eliminated to be on path, given the use of behavioral strategies in our construction and the common form of forward induction arguments.[20] 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.

4.2 Admissibility and iterated dominance

It is well known that sequential equilibria fail admissibility; that is, players may play a weakly dominated strategy on the path of play of a sequential equilibrium. The reason is that sequential optimality is required only at the limit (or equivalently, under the characterization in Proposition 2.1, only ε_n-optimality is required along the sequence of strategy profiles, for some $\varepsilon_n \to 0$). (By contrast, perfect equilibria, where $\varepsilon_n = 0$ for all n, satisfy admissibility.) It is not difficult to see that sequentially stable outcomes may fail admissibility for the same reason. This is unsurprising, since requiring both admissibility and iterated (strict) dominance leads to the non-existence of equilibrium concepts that are not set-valued.[21] We now describe a sense in which admissibility is fragile to payoff perturbations, but sequential stability is not.

Consider the game in Figure 2(b). This game has a unique outcome that is the outcome of a KM-stable set of equilibria, in which player 1 chooses B_1 for sure, as B_1 is its only admissible strategy. Intuitively, any small tremble by player 2 brings player 1's payoff from choosing T_1 below

[19] Note that 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.

[20] Cho (1987) defines a refinement of sequential equilibria, called *forward induction equilibria*, 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.

[21] Kohlberg and Mertens (1986) show that KM-stable sets satisfy admissibility (i.e., only contain equilibria in which players do not play weakly dominated strategies). Also, in Section 2.7.B, they exhibit a game (called Ω) that shows why requiring admissibility and 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 together with admissibility leads to the same non-existence result.

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 ξ, 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.

By contrast, the proposition below shows that sequential stability is robust in the following sense: given a sequentially stable outcome ω, for any perturbation of the game (in terms of trembles), there is a game with nearby payoffs that has an equilibrium outcome close to ω.

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

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

4.3 Sequential stability in subgames

Selten (1965) introduced the concept of subgame-perfect (Nash) equilibria 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 equilibria) and a requirement in others (e.g., sequential equilibria). As well as adding plausibility, subgame perfection facilitates the study of games by enabling the use of backward induction. For example, by iteratively replacing subgames with 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 robustness 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 outcomes of sequential equilibria, which are themselves subgame-perfect. Still, subgame perfection cannot be applied directly to the concept of sequential stability, because sequentially stable outcomes do not specify off-path behavior. As Example 3.2 shows, the limit off-path behavior for the sequence of sequential ε_n-equilibria supporting a sequentially stable outcome may depend on the particular vanishing tremble used.

The following proposition establishes three results that help us study sequential stability via subgames. The first says that a subgame with a unique sequential outcome can be replaced by that outcome without altering the set of sequentially stable outcomes. This lets us iteratively reduce the complexity of a game. The second result says that if an outcome ω is sequentially stable in the game 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 helps us find sequentially stable outcomes by iteratively replacing subgames by their sequentially stable outcomes. Finally, the third result says that the conditional distribution induced by a sequentially stable outcome in an on-path subgame is itself a sequentially stable outcome of the subgame. This provides a way to 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 narrow down the possible on-path behaviors of sequentially stable outcomes.

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.*[22]

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 sequential ε-equilibrium or sequentially stable outcome) and 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-form 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 equilibria), sequential stability is not invariant to changes in the game tree that preserve the reduced normal form of the game.

While there is disagreement on the desirability of invariance as a requirement for equilibrium concepts, most authors agree that invariance to interchanging simultaneous moves is a basic and

[22] By "the game where G' is replaced by ω'" we mean the game in which, at the node where G' is initiated, nature chooses each terminal history z' in the support of ω' with probability $\omega'(z')$. Note that Govindan (1996) proves a result similar to Proposition 4.3(2), but for KM-stable sets of equilibria.

necessary requirement (which is also satisfied by KM-stable sets). Indeed, while a modeler of a particular economic activity may identify the sequence of moves to determine the game tree, she has freedom in how to encode simultaneous moves. The following proposition states that the set of sequentially stable outcomes does not depend on the order of the 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.*

Example 4.3. Although sequential stability does not satisfy invariance to coalescing consecutive moves, 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 3 with $x\in(1,2)$, which Kohlberg and Mertens (1986) use to exemplify the invariance of KM-stability (see their Figures 2 and 3). Note that game (b) is obtained by "coalescing" 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 sequential and a perfect 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 KM-stable in either game. By the same logic, one can show that ω_1 is not a sequentially stable outcome of game (a) or game (b) (note that B_1 is a strictly dominated action in both games), so the only sequentially stable outcome assigns probability one to $(\mathrm{M}_1,\mathrm{T}_2)$. On the other hand, when $x\in(0,1)$, Example 4.1 explains that ω_1 is sequentially stable in (a) but not in (b).

4.5 Approaches to obtaining sequentially stable outcomes

We now discuss procedures for identifying sequentially stable outcomes in extensive-form games without explicitly showing sequential stability for all vanishing trembles (as in Example 3.2). These procedures overcome some of the complications that make using KM-stable sets difficult in practice, as described in the introduction.

Through applying necessary conditions: The first procedure consists in eliminating all candidates for sequentially stable outcomes except one, by applying 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 are sometimes difficult to compute, candidates for sequentially stable outcomes may be drawn from outcomes of a weaker class of 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 17) 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.[23]

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 be difficult to find the right vanishing tremble and then prove that ω is the unique limit equilibrium outcome along it.[24,25] Examples 3.4 and 5.1 illustrate 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 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

[23]Note that a similar argument is difficult to make when using iterated dominance as in Kohlberg and Mertens (1986), as the uniqueness of a KM-stable set in the simpler game only implies that this set is *part* of a KM-stable set of the original game (see Proposition 6 in Kohlberg and Mertens, 1986, which states, "A [KM-]stable set contains a [KM-]stable set of any game obtained by deletion of a dominated strategy").

[24]In Lemma A.1 in the appendix, we provide a convenient characterization: An outcome $\omega\in\Omega$ is sequentially stable if and only if the property in Definition 3.1 holds for some subsequence (ξ_{k_n}) of (ξ_n) (instead of the whole sequence).

[25]In a companion paper (Dilmé, 2023b), 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.

limit equilibrium outcome, which is then the unique sequentially stable outcome. In signaling games à la Spence (1973), for example, such an outcome is often—but not necessarily (see Example 5.1)—the least costly fully separating outcome, called the *Riley outcome*.

Combining techniques: Note that the two procedures described above 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.

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 (Cho and Kreps, 1987) and divinity and universal divinity (Banks and Sobel, 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 set of types who can send message m.[26] Abusing notation, we let $u_\theta(m,r)$ and $u_{\mathrm{r}}(\theta,m,r)$ denote the payoffs of the sender and the receiver, respectively, at $(\theta,m,r) \in Z$, and we let $u_\theta(\omega)$ denote the sender's payoff under outcome ω conditional on the realized type being θ. We let $\mathrm{BR}_m(\mu_m) \subset \Delta(R_m)$ be the set of (mixed) best responses of the receiver to message m when her belief about the sender's type is $\mu_m \in \Delta(\Theta_m)$, and $\mathrm{BR}_m := \cup_{\mu_m \in \Delta(\Theta_m)} \mathrm{BR}_m(\mu_m)$.

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

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

Proposition 5.1. *The outcome 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 $\rho \in \mathrm{BR}_m(\alpha\,\mu_m + (1-\alpha)\,\mu'_m)$ with the following properties: $u_\theta(m,\rho) \le u_\theta(\omega)$ 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 $\mu'_m(\theta) > 0$.*

Banks and Sobel (1987) and Cho and Kreps (1987) find that, in a signaling game with generic payoffs, an outcome satisfies the conditions in Proposition 5.1 (or, more precisely, it satisfies similar but slightly more complicated conditions) if and only if it is KM-stable; see their Theorem 3 and Proposition 4, respectively.[27] Consequently, the following is an immediate corollary of our Propositions 3.2 and 5.1.

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

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

Proposition 5.2. *If an outcome is the unique sequentially stable outcome of G^{sig}, then it is its unique KM-stable outcome.*

To prove Proposition 5.2, we first show that in a signaling game, if ω is extensive-form stable, then it is KM-stable.[28] We then show that if ω is KM-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 3.3) and hence KM-stable, and (ii) there is no other KM-stable outcome, since any KM-stable outcome would be sequentially stable.

5.2 Signaling refinements

Cho and Kreps (1987) proposed several distinct criteria for selecting equilibria in signaling games. These criteria have shaped the handling of equilibrium multiplicity across diverse applications. In this section, we argue that sequential stability is stronger than all of these criteria.

[27] We say an outcome ω is *KM-stable* if, for any vanishing tremble of the reduced normal form of the game, there is a corresponding sequence of Nash equilibrium outcomes converging to ω. KM-stable outcomes exist in games with generic payoffs (see Footnote 3).

[28] This is non-trivial to prove, because the trembles of the reduced-form game (used to determine KM-stability) affect all sender types equally and may be correlated across types, while the sizes of the behavioral trembles (used to determine extensive-form stability) may depend on the type and are uncorrelated across types.

For conciseness, we refer to the IC (Intuitive Criterion), D1, D2, and NWBR_{CK} (i.e., Cho and Kreps's version of NWBR; see below) as the *standard selection criteria* (for signaling games).[29] They are based on the following procedure: First, fix an outcome of a sequential equilibrium. Then, for each off-path message, prune out all types deemed implausible according to the criterion. Finally, if not all types have been pruned out, check whether there is a sequential equilibrium in which, 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. Cho and Kreps (1987) provide intuition and motivation for each selection criterion.

The proposition below shows that sequentially stable outcomes pass the standard selection criteria. An implication is that, for each standard selection criterion, there is an outcome passing it. Additionally, if there is a unique outcome passing one of the standard selection criteria, such an outcome is the unique sequentially stable (and KM-stable) outcome of G^{sig}.

Proposition 5.3. *Let ω be a sequentially stable outcome of G^{sig} and m an off-path message. For each type $\theta \in \Theta_m$, define conditions* IC*,* D1*,* D2*, and* $\text{NWBR}_{\textbf{CK}}$ *as follows:*

IC: $\forall \rho \in \text{BR}_m \;\; u_\theta(m,\rho) < u_\theta(\omega)$.

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

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

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

For each $X \in \{\text{IC},\text{D1},\text{D2},\text{NWBR}_{\textbf{CK}}\}$, let $\hat{\Theta}_X$ be the set of all $\theta \in \Theta_m$ satisfying condition X. Then, if $\hat{\Theta}_X \neq \Theta_m$, there is a sequential equilibrium (σ,μ) with outcome ω where $\mu_m(\hat{\Theta}_X) = 0$.

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 2.1, 3.1–3.3, and 4.2 is a signaling game, we can use the results of Section 5 to analyze it further. See Dilmé (2023b) for more examples.

[29] We omit the criteria of divinity and universal divinity proposed by Banks and Sobel (1987), since they are based on a different methodology. In Dilmé (2023a), we show that sequentially stable outcomes also pass iterated applications of the criteria of Cho and Kreps (1987).

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

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

with $1 < c_{\theta_1} < c_{\theta_0} < 1/\Delta$. Note that the receiver prefers to choose $r = 0$ when $\theta = \theta_0$ and $r = 1$ when $\theta = \theta_1$, while both sender types prefer a low message and want the receiver to choose $r = 1$. Let $\overline{m}$ be the smallest message bigger than $1/c_{\theta_0}$. We consider the following vanishing tremble (ξ_n): Sender types θ_0 and θ_1 tremble to all messages with likelihoods n^{-2} and n^{-1}, respectively (i.e., $\xi_n(m|\theta_0) = n^{-2}$ and $\xi_n(m|\theta_1) = n^{-1}$ for all m), while the receiver trembles to all responses with probability n^{-1} (we initialize n so that all trembles are smaller than 1).

Let $(\varepsilon_n) \to 0$ and let $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n))$ be such that (σ_n) supports an assessment (σ, μ) (which by Proposition 2.1 is a sequential equilibrium). Let $m_+ < 1$ be the highest effort such that $\sigma_n(m|\theta_0) > \xi_n(m|\theta_0)$ for an infinite number of $n \in \mathbb{N}$. It must then be that $\sigma_n(r = 1|m) \to 1$ for all $m > m_+$. 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 type θ_0 prefers not to deviate to choose $m_+ + \Delta$, that is,

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

Since m_+ is optimal for type θ_0, the usual single-crossing property implies that $\sigma_n(m|\theta_1) = \xi_n(m|\theta_1)$ for all $m < m_+$ if n is large enough.[30] Since type θ_0 has to assign positive probability to at least one effort below m_+ in the limit, and since the receiver chooses $r = 0$ if her posterior about the type being θ_1 is small enough, type θ_0 must choose $m = 0$ with positive probability in the limit. Hence,

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

Generically in c_{θ_1}, there are then two cases.[31] In the first case, that is, when $\lfloor (c_{\theta_0}\,\Delta)^{-1} \rfloor\, c_{\theta_0}\,\Delta <$

[30]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 $\sigma_n(m|\theta_1) = \xi_n(m|\theta_1)$ if n is large enough.

[31]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, because of both the structure of the message space and the payoffs (5.1). So KM-stable outcomes cannot be assumed to exist.

$1-c_{\theta_1}\Delta$, type θ_1 strictly prefers $m_+ + \Delta$ to m_+. This implies that σ is uniquely determined by[32]

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

Hence the only candidate for a sequentially stable outcome is the Riley outcome, in which type θ_0 chooses the least costly message, while type θ_1 chooses the cheapest message that allows full separation. Since the Riley outcome is the unique limit equilibrium outcome for the vanishing tremble under consideration, it must be both sequentially stable and KM-stable. In the second case, when $\lfloor (c_{\theta_0}\Delta)^{-1} \rfloor c_{\theta_0}\Delta > 1-c_{\theta_1}\Delta$, type θ_1 strictly prefers m_+ to $m_+ + \Delta$. Thus 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 case, there is again a unique candidate for a sequentially stable outcome, but it does not coincide with the Riley outcome. In this outcome, type θ_0 randomizes between the lowest message and a separating message, while type θ_1 chooses the separating message with probability one. Again, this outcome is both sequentially stable and KM-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, $M := \{m_0 = 0, m_1 = 1\}$, and actions in a grid, $R := \{0, 1/\overline{r}, \ldots, 1-1/\overline{r}, 1\}$, 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 $-(r-\theta)^2$;

---

[32] An example of a sequence (σ_n) converging to σ such that $\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n)$ for all n and some sequence $(\varepsilon_n) \to 0$ is

$$\big(\sigma_n(m|\theta_0), \sigma_n(m|\theta_1), \sigma_n(r=1|m)\big) := \begin{cases} (K_n^0, n^{-1}, n^{-1}) & \text{if } m=0, \\ (n^{-1}, n^{-1}, c_{\theta_0} m) & \text{if } 0<m\leq m_+, \\ (n^{-2}, K_n^1, 1-n^{-1}) & \text{if } m=m_+ + \Delta, \\ (n^{-2}, n^{-1}, 1-n^{-1}) & \text{if } m>m_+, \end{cases}$$

where the choice of each K_n^i ensures that $\sigma_n(\cdot|\theta_i)$ is a probability distribution (note that $K_n^i \to 1$ as $n \to \infty$).

that is, he "tries to match" the belief about type θ_1. We also assume that message m_1 is costly, that type θ_0 prefers high actions, and that type θ_1 prefers intermediate actions:

$$u_{\theta_0}(m,r):=1_{[1/3,1]}(r)-m\,c_{\theta_0} \quad \text{and} \quad u_{\theta_1}(m,r):=1_{[1/4,3/4]}(r)-m\,c_{\theta_1}\ ,$$

where $c_{\theta_0}, c_{\theta_1} \in (0,1)$. Consider an outcome in which both types choose m_1 and the receiver chooses $r=1/2$. Such an outcome passes all standard selection criteria (IC, D1, D2, and NWBR$_{\text{CK}}$), because the sets of receiver 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 in which type θ_1 trembles to m_0 with a higher likelihood than type θ_0, say $\xi_n(m_0|\theta_1):=(n+1)^{-1}$ and $\xi_n(m_0|\theta_0):=(n+1)^{-2}$. Assume there are two sequences (ε_n) and (σ_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 her payoff from choosing m_1, since otherwise, the receiver would assign an increasingly high posterior to type θ_1 after m_0, leading her to choose $r=1$ and making type θ_0 strictly willing to deviate. As a result, the probability with which the receiver plays an action in $[1/3,1]$ after m_0 must tend to $1-c_{\theta_0}$ 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 high probability on $\{k/\overline{r},(k+1)/\overline{r}\}$ after m_0 as n increases. Hence, type θ_1's payoff gain from choosing m_0 instead of m_1 remains positive and bounded away from 0 as n increases, since by doing so, she obtains approximately 1 instead of approximately $1-c_{\theta_1}$. This is a contradiction. It is not difficult to see that the game has a unique sequentially stable outcome in which both types choose m_0 for sure. By Proposition 5.2, such an outcome is KM-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 shares many desirable properties with KM-stability. First, it gives robust predictions: Any perturbation of a game has almost-optimal behavior close to a sequentially stable outcome. Second, a sequentially stable outcome satisfies various plausibility 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 pass most 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 identified by ruling out the alternatives through some vanishing tremble, using properties such as NWBR 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 KM-stable outcomes when they are unique or when payoffs are generic.

Several questions not addressed in our analysis may constitute avenues for future research. First, an axiomatic characterization of sequential stability would be desirable.[33] Second, it may be interesting to investigate which classes of games beyond signaling games feature generic equivalence between KM-stability and sequential stability. Finally, it may be possible to strengthen sequential stability to a criterion that retains the existence properties of sequential stability, yet selects unique outcomes.

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. (2023a): "Iterated exclusion of implausible types in signaling games," Mimeo.

[33] This kind of characterization has proven to be elusive for KM-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.

——— (2023b): "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.

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.

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 in the main text, we state and prove a result that will be useful for proving that an outcome is sequentially stable. It establishes that sequential stability can be equivalently defined in an apparently weaker form than our Definition 3.1.

Lemma A.1. *An outcome $\omega \in \Omega$ is sequentially stable if and only if for any vanishing tremble (ξ_n) there exist a strictly increasing sequence of indexes (k_n) and two sequences $(\varepsilon_n) \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_{k_n}))$ such that (ω^{σ_n}) converges to ω.*

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

$$d(\omega, \Omega^*_\varepsilon(\xi_{k_n})) := \inf_{\omega' \in \Omega^*_\varepsilon(\xi_{k_n})} d(\omega, \omega') \ .$$

(Recall that, as explained in Footnote 8, we use the sup distance between outcomes; that is, for any pair of outcomes ω and ω', $d(\omega, \omega') := \max_{z \in Z} |\omega(z) - \omega'(z)|$). This contradicts our original assumed property of ω, since the vanishing tremble (ξ_{k_n}) does not have a subsequence and corresponding sequences of epsilons and sequential epsilon-equilibria with outcomes converging to ω. $\square$

A.2 Proofs of the results in Section 2

Proof of Proposition 2.1

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

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

(We initialize the index n so that the conditions in Definition 2.1 are satisfied for all n.) It is clear that $\xi_n(a)\to 0$ 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 also that, given our definition of 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 $\sigma_n\in\Sigma^*_{\varepsilon_n}(\xi_n)$. Also, for all $a\in A_*$, we have that $(*)$ tends to 0 as $n\to+\infty$, because $u(a'|\sigma_n)\to u(a'|\sigma,\mu)$ for all $a'\in A$ and also because $u(a|\sigma,\mu)=\max_{a'\in A^{I^a}} u(a'|\sigma,\mu)$ 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 vanishing tremble (ξ_n), a sequence $(\varepsilon_n)\to 0$, and a sequence $(\sigma_n\in\Sigma^*_{\varepsilon_n}(\xi_n))$ such that $\omega^{\sigma_n}\to\omega$. Let (k_n) be strictly increasing and satisfy that (σ_{k_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 $\sigma_n\in\Sigma^*_{\varepsilon_n}(\xi_n)$), 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 (σ_{k_n}), so (σ,μ) is a sequential equilibrium with outcome ω. □

A.3 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 $G(\xi)$ has a sequential ε-equilibrium with outcome ε'-close to ω for all $\|\xi\|<\delta_{\varepsilon,\varepsilon'}$. Then, there exists a vanishing tremble (ξ_n) such that there is no $(\sigma_n\in\Sigma^*_\varepsilon(\xi_n))$ 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 $\|\xi\|<\delta_{\varepsilon,\varepsilon'}$, then $G(\xi)$ has a sequential ε-equilibrium with outcome at a distance lower than ε' from ω. Take a vanishing tremble (ξ_n). Let $\overline{u}:=\max_{i\in N}(\max_{z\in Z} u_i(z)-\min_{z\in Z} u_i(z))$. Fix some sequence $(\hat{\varepsilon}_n)$ strictly decreasing towards 0 with $\hat{\varepsilon}_0=\overline{u}$, and recursively define each ε_n as follows:

1. We define $n_0:=0$ and $\varepsilon_0:=\hat{\varepsilon}_0$.
2. For all $k\geq 1$ we let $n_k:=\min\{n>n_{k-1}\,|\,\|\xi_{n'}\|<\delta_{\hat{\varepsilon}_k,\hat{\varepsilon}_k}$ 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 $(\varepsilon_n)\to 0$. Note that, for each k, $\|\xi_n\|<\delta_{\hat{\varepsilon}_{k-1},\hat{\varepsilon}_{k-1}}$ for all $n\in\{n_{k-1}+1,\dots,n_k\}$. Hence, there exists a sequence (σ_n) where, for each k and $n\in\{n_{k-1}+1,\dots,n_k\}$, σ_n is a sequential $\hat{\varepsilon}_{k-1}$-

equilibrium of $G(\xi_n)$ with outcome $\hat{\varepsilon}_{k-1}$-close to ω. Hence, since $\hat{\varepsilon}_{k-1} = \varepsilon_n$ for all $n \in \{n_{k-1}+1, ..., n_k\}$, we have that $\omega^{\sigma_n} \to \omega$. Since the argument holds for any vanishing tremble (ξ_n), ω is sequentially stable. □

Proof of Corollary 3.1

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

Proof of Proposition 3.2

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

Proof of Proposition 3.4

Proof. It is convenient to first prove Proposition 3.4 and then Proposition 3.3. Let $\hat{G}$ be the agent extensive form of G.[34] 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 $\tilde{u} \equiv (\tilde{u}_i : Z \to \mathbb{R})_{i \in \hat{N}}$, we let $\hat{G}(\tilde{u})$ be the agent extensive form of G with payoff function given by $\tilde{u}$ instead of $\hat{u}$. Let $(\hat{u}_k)$ be a sequence of payoff functions converging to $\hat{u}$ such that, for each k, $\hat{G}(\hat{u}_k)$ has an extensive-form stable outcome denoted ω_k, which by Proposition 3.2 is also sequentially stable.[35] Note that, since an extensive-form stable outcome exists for generic payoff functions (recall Footnote 3), a sequence $(\hat{u}_k)$ with the previous properties exists.[36] We can assume without loss of generality for our argument that (ω_k) converges to some outcome ω. Then, the following lemma shows that ω is sequentially stable.

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

---

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

[35]Recall that, as explained in Footnote 8, we use the sup distance between payoff functions; that is, for any pair of payoff functions u and u', $d(u,u') := \max_{i \in N} \max_{z \in Z} |u_i(z) - u_i'(z)|$.

[36]Because $\hat{G}$ coincides with its agent extensive form, Kohlberg and Mertens (1986)'s definition of KM-stability is based on perturbing the game as follows. Fix a vector $(\delta_i)_{i \in N} \in (0,1)^N$ and a completely mixed strategy profile $\hat{\sigma} \in \Sigma$. Then, when players use a strategy profile σ in the perturbed game, the corresponding outcome and payoffs are computed by replacing each strategy σ_i by $(1-\delta_i)\sigma_i + \delta_i \hat{\sigma}_i$ in the unperturbed game. Defining $\xi_i := \delta_i \hat{\sigma}_i$, their formulation becomes equivalent to the tremble-based formulation we use, where agents choose strategies σ_i satisfying $\sigma_i \geq \xi_i$.

Proof. Let (u_k) be a sequence of payoff functions converging to u. Let $(\omega_k)\to\omega$ be such that each ω_k is sequentially stable in $G(u_k)$. Fix a vanishing tremble (ξ_n). For each $k\in\mathbb{N}$, let $(\varepsilon_{k,n})\to 0$ and $(\sigma_{k,n}\in\Sigma^*_{\varepsilon_{k,n}}(\xi_n, u_k))$ be such that $\omega^{\sigma_{k,n}}\to\omega_k$ as $n\to\infty$, which exist by the assumption that ω_k is sequentially stable in $G(u_k)$. Note that for each $a\in A$ and $k,n\in\mathbb{N}$ with $\sigma_{k,n}(a)>\xi_n(a)$, we have

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

Hence,

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

Let (n_k) be a sequence of indexes such that $\varepsilon_{k,n_k}\to 0$ and $\omega^{\sigma_{k,n_k}}\to\omega$ as $k\to\infty$, which exists by a standard diagonal argument.[37] Then, defining $\varepsilon_k:=2\,d(u,u_k)+\varepsilon_{k,n_k}$, we have that each $\sigma_k:=\sigma_{k,n_k}$ is a sequential ε_k-equilibrium of $G(\xi_{n_k})$. Then, since $\varepsilon_k\to 0$ as $k\to\infty$, and since the argument holds for any vanishing tremble (ξ_n), Lemma A.1 implies that ω is sequentially stable. □

□

Proof of Proposition 3.3

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 3.2), 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 extensive-form stable outcome. Let $(\hat{\xi}_n)$ be a tremble such that there is no sequence of indexes (k_n) such that there is some sequence $(\hat{\sigma}_n\in\Sigma^*_0(\hat{\xi}_{k_n}))$ with outcomes converging to ω (where $(\hat{\xi}_n)$ exists since ω is not extensive-form stable). By Proposition 3.4, each $\hat{G}(\hat{\xi}_{\hat{n}})$ has at least one sequentially stable outcome $\omega_{\hat{n}}$. We let (k_n) be a sequence of indexes such that ω_{k_n} converges to some ω'. Since, for each n, $\omega^{\hat{\sigma}_n}=\omega_{k_n}$ for some $\hat{\sigma}_n\in\Sigma^*_0(\hat{\xi}_{k_n})$, the previous assumption on $(\hat{\xi}_n)$ 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 below.

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

Lemma A.3. *Let (ξ_n) be a vanishing tremble. Let $(\omega_n) \to \omega$ be such that each ω_n is sequentially stable in $G(\xi_n)$. Then, ω is a sequentially stable outcome of G.*

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

□

A.4 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 $A' \subset A \backslash \{\hat{a}\}$ be its set of actions. Let (ξ'_n) be a vanishing tremble in G', and let $\underline{\xi}'_n := \min\{\xi'_n(a') | a' \in A'\}$. Define the vanishing tremble (ξ_n) as follows:

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

for all $a \in A$, and note that (ξ_n) is a vanishing tremble in G. Note also that, under the vanishing tremble (ξ_n), any history in H not belonging to H' (i.e., with some $a \notin A'$) has a vanishing relative likelihood with respect to any history in H'. Let $(\varepsilon_n) \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n))$ 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 (σ_n) supports some assessment (σ, μ), which by Proposition 2.1 is a sequential equilibrium. Note that (σ, μ) has outcome ω.[38] Note also that, if n is large enough, it must be that $\sigma_n(\hat{a}) = \xi_n(\hat{a})$, since by assumption $\hat{a}$ is not sequentially optimal under (σ, μ), hence there is some $\hat{a}' \in A^{I^a}$ such that

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

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

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

[38] Recall that, by Lemma A.1, it is enough to prove that the property in Definition 3.1 holds for a subsequence of (ξ_n).

Note that $\sigma'_n \in \Sigma'(\xi'_n)$ (i.e., σ'_n belongs to the set of strategy profiles of G' satisfying $\sigma'_n \geq \xi'_n$). We claim that there is some sequence $(\varepsilon'_n) \to 0$ such that $(\sigma'_n \in \Sigma'^{*}_{\varepsilon'_n}(\xi'_n))$. This follows from the fact that all information sets of G that contain both histories in H' and not in H', the relative weight of histories not in H' shrinks to 0 as n increases because 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 $n \to \infty$.[39] □

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 parts 1 and 2 are equivalent 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 $\|\xi\| < \delta$, then there is some u' with $\|u' - u\| < \varepsilon$ such that $G(\xi, u')$ 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 $\|\xi\| < \delta$, then $G(\xi)$ has a sequential ε-equilibrium with outcome ε'-close to ω. Applying Proposition 3.1, we then have that ω is sequentially stable.

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 $\|\xi\| < \delta$, then $G(\xi)$ has a sequential ε-equilibrium with outcome ε'-close to ω. We fix some ξ with $\|\xi\| < \delta$ and let $\sigma \in \Sigma^{*}_{\varepsilon}(\xi)$ be a sequential 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 $G(\xi, u')$.

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

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 so that ε-optimal actions under u will become exactly optimal under u'. To do so, recall that $Z^a \subset Z$ is the set of terminal histories containing $a \in A$. We denote the information sets $\mathcal{I} := \{I_1, \ldots, I_{|\mathcal{I}|}\}$. 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 $j = 1, \ldots, |\mathcal{I}|$, 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 $\hat{u}^{j-1}(a|\sigma) \geq \overline{u}_{\iota(I_j)} - \varepsilon$. Note that $a \in A^{I_j}$ is such that $\sigma(a) > \xi(a)$ only if $a \in A^{I_j}_*$.
2. We define $\hat{u}^j$ as a payoff function assigning to each $i \in N$ and $z \in Z$ the value $\hat{u}^{j-1}_i(z)$, except for the value assigned to player $\iota(I_j)$ at terminal histories $z \in Z^a$ for some $a \in A^{I_j}_*$, where

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

where $K(I_j)$ is chosen such that

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

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

$$\hat{u}^j_{\iota(I_j)}(z) - \hat{u}^j_{\iota(I_j)}(z') = \hat{u}^{j-1}_{\iota(I_j)}(z) - \hat{u}^{j-1}_{\iota(I_j)}(z')$$

for all $z, z' \in Z^a$, 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 precedes 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 $\hat{u}^{|\mathcal{I}|}$. $\square$

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 22). Let A' be the set of actions of G', $Z'_{\omega'}$ 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 sequentially stable outcomes. We divide the proof into two subparts.

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

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

for all $a \in A$. Let $(\varepsilon_n) \to 0$ and $(\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n))$ be such that $\omega^{\sigma_n} \to \omega$ (which exist because ω is sequentially stable in G). By Proposition 2.1 and the fact that ω' is the unique sequential outcome in G', we have that the conditional distribution of ω^{σ_n} on Z' 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 $\hat{\sigma}_n(z') := \omega'(z')$ for all terminal histories $z' \in Z'_{\omega'}$ (note that, in G', $A'^{I'} = Z'_{\omega'}$, and so nature plays each $z' \in A'^{I'}$ with probability $\omega'(z') > 0$). It is clear that $\omega^{\hat{\sigma}_n} \to \omega$. Also, it is easy to see that there exists some $(\hat{\varepsilon}_n) \to 0$ such that $\hat{\sigma}_n \in \hat{\Sigma}^*_{\hat{\varepsilon}_n}(\hat{\xi}_n)$ 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 (ξ_n) in G and let (ξ'_n) be its restriction to G'. Let $(\hat{\xi}_n)$ be a vanishing tremble in $\hat{G}$ satisfying that $\hat{\xi}_n(a) := \xi_n(a)$ for all $a \in \hat{A}$ and $\hat{\xi}_n(z') \le \omega'(z')$ for all $z' \in Z'_{\omega'}$. Let $(\hat{\varepsilon}_n) \to 0$ and $(\hat{\sigma}_n \in \hat{\Sigma}^*_{\varepsilon'_n}(\hat{\xi}_n))$ be such that $\omega^{\hat{\sigma}_n} \to \hat{\omega}$ (which exist because $\hat{\omega}$ is sequentially stable in $\hat{G}$). Let $(\varepsilon'_n) \to 0$ and $(\sigma'_n \in \Sigma'^*_{\varepsilon'_n}(\xi'_n))$ be such that $\omega^{\sigma'_n} \to \omega'$ (which exist by Proposition 2.1 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 $(\varepsilon_n) \to 0$ such that $\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n)$ 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 Z' be the set of terminal histories of G'.[40] Assume, for contradiction, that the restriction of ω to

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

G', $\omega' := \omega|_{Z'}$, is not a sequentially stable outcome of G', and let (ξ'_n) be a tremble sequence of G' destroying it (i.e., such that there are no corresponding sequences (ε'_n) and (σ'_n) with the properties in Definition 3.1). Then, it is clear that any perturbation ξ_n that coincides with ξ'_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 $(a_1, ..., a_J) \in Z$, 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 Z by $\mathcal{T}(Z)$, 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 (ξ_n), and let (ε_n) and (σ_n) satisfy the conditions in Definition 3.1, with $\omega^{\sigma_n} \to \omega$ (which exist since ω is sequentially stable). We argue that (ξ_n), (ε_n) and (σ_n) also satisfy the conditions in Definition 3.1 in G' and the limit of (ω^{σ_n}) is ω'.[41] To see this, note that for a given $a \in A$, player $\iota(I^a)$'s payoff from playing a in G under σ_n is

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

where the second equality follows because of $\Pr^{\sigma_n}(\mathcal{T}(z)) = \Pr^{\sigma_n}(z)$ (since z and $\mathcal{T}(z)$ contain the same actions), $u'_{\iota(I^a)}(\mathcal{T}(z)) = u_{\iota(I^a)}(z)$ (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 (ξ_n), (ε_n), and (σ_n) also satisfy the conditions in Definition 3.1 in G', hence ω' is a sequentially stable outcome of G'. □

[41] Note 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 (ξ_n) is also a vanishing tremble of G', and σ_n is a sequential ε_n-equilibrium of $G'(\xi_n)$ for each n.

A.5 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 (ξ_n) such that

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

and such that $\xi_n(m|\theta')/\xi_n(m|\theta) = \mu_m(\theta')/\mu_m(\theta)$ for all $\theta, \theta' \in \Theta_m$ with $\theta \in \text{supp}(\mu_m)$.[42] Let (ε_n) and (σ_n) be two sequences satisfying that $\varepsilon_n \to 0$, $\sigma_n \in \Sigma^*_{\varepsilon_n}(\xi_n)$ 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 (σ_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 $\sigma_n(m|\theta) = \xi_n(m|\theta)$ for all $n > \overline{n}$. Note further that it must be that $\theta \in \text{supp}(\mu_m)$, since by the definition of (ξ_n), we have that $\xi_n(m|\theta')/\xi_n(m|\theta'') = 0$ 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 $\theta' \in \Theta_m$.[43] Hence, the result holds for $\alpha := 1$, $\mu'_m(\cdot) := \mu''(\cdot|m)$, and $\rho := \sigma(\cdot|m)$.

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

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

---

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

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

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 $\sigma_n(\theta') = \xi_n(\theta')$ 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 (ξ_n). We will construct a strictly increasing sequence (k_n) and a sequence (σ_{k_n}) such that $\sigma_{k_n} \in \Sigma^*_{\varepsilon_{k_n}}(\xi_n)$ for all n for some sequence $(\varepsilon_{k_n}) \to 0$ and $\omega^{\sigma_{k_n}} \to \omega$ 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 $M_0 := \{m^1, \ldots, m^{|M_0|}\}$ (we assume that $|M_0| \geq 1$ since the result is trivial otherwise). We first construct $\sigma_n(m|\theta)$ and $\sigma_n(r|m)$ for all $m \in M_0$, $\theta \in \Theta_m$, and $r \in R_m$. We do it by first proceeding recursively over the set of messages that are off-path under ω, and then we will define the values for on-path messages. We begin with $k=1$ and $(j_n^0) := (n)$. Then, for each $k = 1, \ldots, |M_0|$, we proceed as follows:

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

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

4. We finally define $\sigma_n(r|m^k)$ as
$$\sigma_n(r|m^k) := \begin{cases} \xi_n(r|m^k) & \text{if } \rho(r) = 0, \\ \rho(r)\big(1 - \sum_{r'\notin \text{supp}(\rho)} \xi_n(r'|m^k)\big) & \text{if } \rho(r) > 0. \end{cases} \tag{A.3}$$
   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

$$\sigma_n(m|\theta) := \begin{cases} \xi_n(m|\theta) & \text{if } \omega(m|\theta)=0, \\ \omega(m|\theta)\left(1-\sum_{\theta'|\omega(m|\theta')=0} \xi_n(m|\theta')\right) & \text{if } \omega(m|\theta)>0, \end{cases}$$

where $\omega(m|\theta)$ is the probability that type θ sends m under ω, and also

$$\sigma_n(r|m) := \begin{cases} \xi_n(r|m) & \text{if } \omega(r|m)=0, \\ \omega(r|m)\left(1-\sum_{r'|\omega(r'|m)=0} \xi_n(r'|m)\right) & \text{if } \omega(r|m)>0, \end{cases}$$

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 $\sigma_n \in \Sigma(\xi_n)$, that $\omega^{\sigma_{j_n}} \to \omega$, and that there is some sequence $(\varepsilon_n) \searrow 0$ such that $\sigma_{j_n} \in \Sigma^*_{\varepsilon_{j_n}}(\xi_{j_n})$ 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 KM-stable outcome, we need some notation regarding the normal form of the extensive-form 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 $(\hat{\xi}_n) \equiv ((\hat{\xi}^i_n)_{i \in N})$, where $\hat{\xi}^i_n : \vec{A}^i \to (0,1]$ is such that $\hat{\xi}^i_n(\vec{a}^i) \to 0$ as $n \to \infty$ for all $\vec{a}^i \in \vec{A}^i$. Then, ω is *stable* if, for any normal-form vanishing tremble $(\hat{\xi}_n)$, there is a sequence $(\hat{\sigma}_n)$ such that (i) $\hat{\sigma}^i_n(\vec{a}^i) \geq \hat{\xi}^i_n(\vec{a}^i)$ for all $\vec{a}^i$ and n, (ii) $\hat{\sigma}^i_n(\vec{a}^i) > \vec{\xi}^i_n(\vec{a}^i)$ 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 KM-stable. Let ω be an extensive-form stable in G^{sig}. Let $(\hat{\xi}_n)$ be a normal-form vanishing tremble. Define the following extensive-form vanishing tremble for all a and n:

$$\xi_n(a) := \sum_{\vec{a} \ni a} \hat{\xi}_n^i(\vec{a}^i) \; .$$

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

$$\sum_{a \in A^I} \xi_n(a) = \sum_{a \in A^I} \sum_{\vec{a}^i \ni a} \hat{\xi}_n^i(\vec{a}^i) = \sum_{\vec{a}^i \in \vec{A}^i} \hat{\xi}_n^i(\vec{a}^i) = \overline{\xi}^i \; .$$

Since ω is extensive-form stable, there is a sequence of Nash equilibria ($\sigma_n \in \Sigma_0^*(\xi_n)$) with outcomes converging to ω. Taking a subsequence if necessary, assume (σ_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

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

Note that

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

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

Part 2: Proof that if ω is KM-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 (ξ_n). Let $\xi_n(I) := \sum_{a \in A^I} \xi_n(a)$ be the sum of the trembles of information set I. We define $\overline{\xi}_n := \max_\theta(\xi_n(I_\theta))$. Note that $\overline{\xi}_n \to 0$.

Fix a KM-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

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

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

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

Since ω is KM-stable, there is a sequence $(\hat{\sigma}_n)$ 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 $r \in R_m$:

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

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

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

$$\begin{aligned} \sigma_n(m|\theta) &\geq \sum_{\vec{m}|\vec{m}_\theta = m} \hat{\xi}^{\mathrm{s}}_n(\vec{m}) = \sum_{\vec{m}|\vec{m}_\theta = m} \prod_{\theta' \in \Theta_m} \xi'_n(\vec{m}_{\theta'}|\theta') = \Big(\prod_{\theta' \in \Theta_m \setminus \{\theta\}} \xi'_n(I_{\theta'}) \Big) \xi'_n(m|\theta) \\ &= \big(\overline{\xi}_n^{1/|\Theta|}\big)^{|\Theta|-1}\, \overline{\xi}_n^{1/|\Theta|-1} \xi_n(m|\theta) = \xi_n(m|\theta)\ , \end{aligned}$$

where we used that $\xi'_n(I_\theta) = \overline{\xi}_n^{1/|\Theta|}$ by construction. Hence, since $\sigma_n(r|m) \geq \xi_n(r|m)$, we have that $\sigma_n \in \Sigma(\xi_n)$. Standard continuity arguments imply that, since $\hat{\sigma}_n$ is a Nash equilibrium of the normal-form game perturbed with $\hat{\xi}_n$, σ_n is asymptotically sequentially optimal for all types θ.[44] Then, ω is sequentially stable. $\square$

[44]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 (σ_n) and $(\hat{\sigma}_n)$, and since the belief of the receiver after each message m coincide under both sequences (σ_n) and $(\hat{\sigma}_n)$.

Proof of Proposition 5.3

Proof. We prove the result for NWBR_{CK}, as it is the strongest statement. The other cases 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 $u_\theta(m,\rho)\leq u_\theta(\omega)$ 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 $\rho\in\text{BR}_m$, 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 $\hat{\rho}\in\text{BR}_m(\alpha\,\mu_m+(1-\alpha)\,\hat{\mu}'_m)$ such that the properties in Proposition 5.1 hold. Define $(\check{\sigma}',\check{\mu}')$ as

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

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 $\check{\mu}'_m(\hat{\Theta})=0$. $\square$