Lecture 2: Equilibrium Refinements

Alexander Wolitzky

MIT

14.126, Spring 2024

Overview: Backward Induction

Many games have lots Nash equilibria—especially dynamic games, where many NE are supported by "non-credible threats."

Simplest way to exclude non-credible threats is backward induction (BI).

BI is well-defined in games of perfect information.

- In these games, it can be seen an an implication of "ultra-rationality."
- Formally, BI is an implication of a (carefully crafted) definition of "common knowledge of rationality."
- However, usual BI definition doesn't apply in more general extensive-form games.

Overview: Subgame Perfection

Subgame perfection applies more generally: a strategy profile in an extensive-form game is a *subgame perfect equilibrium* if its restriction to each proper subgame is a Nash equilibrium.

- Recall: a proper subgame is an extensive-form game starting at any node (and not "cutting" any information set).

Subgame perfection reduces to BI in perfect info games, but applies more generally.

However, subgame perfection is still too permissive in most imperfect info games, because there's aren't many proper subgames.

- In incomplete info games, there are no proper subgames.

This issue has led to the development of several important refinements of SPE, which reduce to SPE under perfect info.

- Perfect Bayesian equilibrium, sequential equilibrium, perfect equilibrium, proper equilibrium, . . .

Overview: Backward vs. Forward Induction

Backward induction and its refinements like SPE and sequential equilibrium are based on the idea that other players will be rational in the future, even if they have taken apparently irrational actions in the past.

- Makes sense if off-path past actions are "mistakes" or "trembles."

In contrast, "forward induction" is the idea that other players' past actions should be given a rational interpretation when possible, even when they are off-path.

- Makes sense if off-path past actions are "deliberate attempts to influence opponents' play."
- Motivates another set of refinements, including strategic stability (Kohlberg Mertens 1986).

BI-based refinements are ubiquitous in dynamic games.
FI-based refinements come up less often but are important for signaling games and related models (e.g., cheap talk, bargaining).

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- Perfect equilibrium
- Proper equilibrium
- Forward induction
- Strategic stability
- Robustness of refinements

Extensive-Form Games

A **extensive-form game** consists of

- A finite set of **players** $I = \{1, \ldots, N\}$,
- A set Z of **terminal nodes**, and **vNM utility functions** $u_i : Z \to \mathbb{R}$ for each player.
- A set X of **non-terminal nodes**, where for each node $x \in X$ we have
  - A player $i(x)$ who moves at x.
  - A set of possible **actions** $A(x)$.
  - A **successor node** $n(x, a) \in X \cup Z$ resulting from each possible action.

  Moreover, $X \cup Z$ is a **tree**: there is a unique **initial node** $o \in X$, and every other $x \in X \cup Z$ has exactly one $x' \in X$ and $a' \in A(x')$ s.t. $x = n(x', a')$.
- **Information sets** $h \in H$ that partition X such that if $x, x' \in h$ then $i(x) = i(x')$ and $A(x) = A(x')$.

Strategies in Extensive-Form Games

A **pure strategy** for player i is a function $s_i : H_i \to A_i$ s.t. $s_i(h) \in A(h)\ \forall h \in H_i$.

- A strategy is a complete contingent plan.

A **mixed strategy** is a probability distribution over pure strategies.

- Randomize over plans at the beginning.

A **behavior strategy** is a sequence of probability distributions over actions, i.e. a function $\sigma_i : H_i \to \Delta(A_i)$ s.t. $\operatorname{supp} \sigma_i(h) \subseteq A(h)$ $\forall h \in H_i$.

- Randomize as you go.

Kuhn's theorem: in games with **perfect recall** (see FT Ch. 3 for the definition), mixed and behavior strategies are equivalent, in that for any mixed strategy there is a behavior strategy that gives the same outcome (distribution over terminal nodes) for all opposing strategy profiles, and vice versa.

Normal and Agent-Normal Form

The **normal** (or **strategic**) form of an extensive-form game is the strategic-form game with players I, strategies $S_i = \{s_i : H_i \to A_i\}$, and payoffs $u_i(s) = u_i(z(s))$.

- Game where players choose complete contingent plans.

The **agent-normal** (or **agent-strategic**) form of an extensive-form game is the strategic-form game with players $(i(h))_{h \in H}$, strategies $S_{i(h)} = A(h)$, and payoffs $u_{i(h)}(s) = u_{i(h)}(z(s))$.

- Game where we view "player i choosing at info set h" and "player i choosing at info set h'" as different players.

This distinction matters for thinking about backward and forward induction.

- E.g., is an off-path move due to a tremble in the normal form (so all bets are off regarding future play) or a tremble in the agent-normal form (so still expect rational play in the future)?

Subgame Perfect Equilibrium

A **subgame** of an extensive-form game (I, X, Z, A, i, n, u, h) is another extensive-form game consisting of

- a subset of nodes $X' \cup Z' \subseteq X \cup Z$ consisting of a single node and all its successors, s.t. if $x \in X'$ and $x, x' \in h$ for some info set h then $x' \in X'$. ("No cutting info sets.")
- The same (I, A, i, n, u, h), restricted to $X' \cup Z'$.

A strategy profile in an extensive-form game is a **subgame perfect equilibrium (SPE)** if its restriction to each subgame is a Nash equilibrium.

Backward induction theorem: every finite extensive-form game of perfect information (i.e., all singleton info sets) has a pure-strategy SPE, which can be found by backward induction. For generic payoffs, the SPE is unique.

However, with imperfect information SPE is too permissive, as there are few or no proper subgames.

Outline

- Extensive-form games and subgame-perfect equilibrium
- **Perfect Bayesian equilibrium**
- Sequential equilibrium
- Perfect equilibrium
- Proper equilibrium
- Forward induction
- Strategic stability
- Robustness of refinements

Beliefs and Assessments in Extensive-Form Games

The best-known extensive-form refinements—perfect Bayesian equilibrium (PBE) and sequential equilibrium—are based on the idea of including players' *beliefs* as well as strategies as part of the definition of an equilibrium.

A **belief system** μ is a mapping $\mu : H \to \Delta(X)$ s.t. $\operatorname{supp} \mu(h) \subseteq h \; \forall h$.

- Interpretation: $\mu(h)$ is $i(h)$'s belief about what node she's at in info set h_i.

An **assessment** (σ, μ) consists of a strategy profile and a belief system.

PBE and sequential equilibrium are *assessments* (not just strategy profiles!) with certain properties.

Weak PBE

There are several definitions of PBE, but they are all refinements of "weak PBE" (also sometimes called "weak sequential equilibrium," but this sounds old-fashioned to me):

Definition

A **weak PBE** is an assessment (σ, μ) such that:

1. σ is **sequentially rational** given μ: σ_i maximizes $u_i\left(\left(\sigma_i', \sigma_{-i}\right)|h, \mu(h)\right)$ $\forall i, h \in H_i$.
2. μ is **consistent with Bayes' rule on-path**: letting $P^{\sigma}(x)$ be the prob of reaching node x under σ, we have $\mu(x|h) = P^{\sigma}(x) / P^{\sigma}(h)$ for all h s.t. $P^{\sigma}(h) > 0$ and all $x \in h$.

Sequential rationality: each player maximizes her expected payoff conditional on reaching **any** info set, given her beliefs and her opponents' strategies.

Bayes' rule on-path: minimal requirement, often too weak...

Problems with Weak PBE; Alternatives

Advantages of weak PBE: easy to define, well-defined in all extensive-form games, sometimes useful.

Disadvantages:

1. Does not always refine SPE (!). (In general, too permissive.)
2. Does not imply Bayesian updating off-path
3. Allows "signaling what you don't know"

FT Ch. 8 give a more refined notion of PBE without these problems. However, it is only well-defined for a class of extensive-form games ("multistage games with observed actions").

- FT's definition ends up being kind of similar to sequential equilibrium (but only for MGOA, while allowing infinite actions), which we turn to next.

There are various other definitions of PBE in the literature. All are in between weak PBE and sequential equilibrium. Applications should be clear about what definition they're using.

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- **Sequential equilibrium**
- Perfect equilibrium
- Proper equilibrium
- Forward induction
- Strategic stability
- Robustness of refinements

Sequential Equilibrium: Overview

Sequential equilibrium (Kreps Wilson 1982) makes stronger predictions by imposing a stronger consistency requirement: off-path beliefs are derived as the limit of Bayes updates from completely mixed strategies.

- "Completely mixed" means positive probability on all actions.
- This requires finite action sets, which is an important limitation of sequential equilibrium.

Sequential Equilibrium: Definition

Definition

A **sequential equilibrium** is an assessment (σ, μ) such that:

1. σ is **sequentially rational** given μ: σ_i maximizes $u_i\left(\left(\sigma'_i, \sigma_{-i}\right)|h, \mu\left(h\right)\right)$ $\forall i, h \in H_i$.
2. (σ, μ) is **(Kreps-Wilson) consistent**: there exists a sequence $\left(\sigma^m, \mu^m\right) \to (\sigma, \mu)$ such that each σ^m is completely mixed and $\mu^m\left(x|h\right) = P^{\sigma^m}\left(x\right) / P^{\sigma^m}\left(h\right)$ for **all** $h, x \in h$.

Notes:

- Optimality not imposed on the σ^m. Can think of difference between σ and σ^m as "trembles."
- Important that σ^m is a strategy profile. Implies that trembles are independent across players and info sets. Restrictive.
- Players beliefs are all derived from the same sequence of trembles: i.e., players agree about distribution of off-path actions. Also restrictive.

Sequential Equilibrium: Properties

SE always exists in finite games.

- Will follow from existence of trembling-hand perfect equilibrium.

SE refines SPE: if (σ, μ) is a SE, then σ is a SPE.

- Follows from sequential rationality and requirement that beliefs match the distribution induced by strategies in each subgame.

The set of SE is UHC with respect to payoffs.

- However, SE is sensitive to the details of the extensive form, which determines which trembles are "allowed."

There is often a continuum of SE, with multiplicity arising from multiplicity of consistent off-path beliefs. But for generic assignments of payoffs to terminal nodes, there are finitely many SE outcomes.

Sequential Equilibrium vs PBE

- SE is defined for all finite games.
  - There are some attempts to generalize it to infinite games, including recently by Myerson and Reny (ECMA 2020).
- Weak PBE is defined for all games, but is often too weak to be useful. Most other PBE notions are defined for more limited classes of games, but do allow infinite actions.
  - There are some attempts to define notions of PBE that are more restrictive than weak PBE but apply beyond MGOA. E.g., Battigalli (JET 1996), Watson (2017), Sugaya and Wolitzky (RESTUD 2021, following Myerson ECMA 1986).
- In games where both are defined, SE is a refinement of PBE.
- FT show that their notion of PBE coincides with SE in 2-stage games (e.g., signaling games) and in games where each player has at most two types.

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- **Perfect equilibrium**
- Proper equilibrium
- Forward induction
- Strategic stability
- Robustness of refinements

(Trembling-Hand) Perfect Equilibrium: Overview

Perfect and proper equilibria are defined for normal-form games but are motivated by similar extensive-form considerations as PBE and SE.

- Definitions don't involve beliefs, but outcomes are similar to those under belief-based refinements like PBE and SE.

A (trembling-hand) perfect equilibrium (Selten 1975) is a Nash equilibrium where each player's strategy remains a best response to **some** perturbation of the opponents' strategies.

Formally, there are three standard (equivalent) definition. . .

Perfect Equilibrium: Definition

Definition

A strategy profile σ is a **perfect equilibrium** if it has any of the following equivalent properties:

1. It is the limit of a sequence σ^m of completely mixed strategy profiles such that $\sigma_i \in \operatorname{argmax}_{\sigma_i'} u_i \left(\sigma_i', \sigma_{-i}^m\right) \forall i, m$.
2. It is the limit of a sequence σ^ε of ε-**constrained equilibria**, where σ^ε is an ε-constrained equilibrium if there exist $\left(\varepsilon\left(s_i\right)\right)_{i,s}$ satisfying $0 < \varepsilon\left(s_i\right) < \varepsilon$ such that σ^ε is a NE in the game where each player i is constrained to take each strategy s_i with probability at least $\varepsilon\left(s_i\right)$.
3. It is the limit of a sequence σ^ε of ε-**perfect equilibria**, where σ^ε is an ε-perfect equilibrium if it is completely mixed but each strategy s_i that is not a best reply to σ_{-i}^ε is played with probability less than ε.

Perfect Equilibrium: Properties

Perfect equilibrium always exists in finite games.

- Apply Nash existence in the ε-constrained game to show existence of ε-constrained equilibrium; take $\varepsilon \to 0$; a convergent subsequence exists by compactness.

Every perfect equilibrium is a NE (by UHC of best responses).
Conversely, any **strict** NE is perfect (as a strict BR remains a BR against perturbations).
Likewise, any **completely mixed** NE is perfect (take $\sigma^m = \sigma \ \forall m$).

Hence, perfect equilibrium is a refinement of NE, and has bite only when a player takes a weak BR against an equilibrium strategy that is not completely mixed.

- However, this situation is typical in dynamic games.

Perfection in Normal vs. Agent-Normal Form

A perfect equilibrium in the normal form of an extensive-form game need not be subgame perfect, because trembles at different nodes can be correlated.

However, if σ is a perfect equilibrium in the agent-normal form of an extensive form game then it is subgame perfect: moreover, there exists μ s.t. (σ, μ) is sequential.

- **Proof.** Take the same σ^m as in the definition of perfection and calculate the limit μ. Perfection requires optimality against each σ^m, which implies optimality at the limit belief.

However, for generic assignments of payoffs to terminal nodes, the sets of perfect and sequential equilibrium outcomes coincide.

- Intuitively, additional requirement of optimality against each σ^m only has bite when there are payoff ties.

Robustness to All Trembles?

A strategy profile σ is **truly perfect** (or **strictly perfect**) if $\sigma_i \in \operatorname{argmax}_{\sigma'_i} u_i \left(\sigma'_i, \sigma^m_{-i}\right)$ $\forall i$ for **all** $\sigma^m \to \sigma$ and all sufficiently large m.

Any strict equilibrium is truly perfect.

However, a truly perfect equilibrium may not exist:

	L	R
U	$3, 2$	$2, 2$
M	$1, 1$	$0, 0$
D	$0, 0$	$1, 1$

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- Perfect equilibrium
- **Proper equilibrium**
- Forward induction
- Strategic stability
- Robustness of refinements

Proper Equilibrium: Definition

Proper equilibrium (Myerson 1978) refines perfect equilibrium by requiring that more costly mistakes are infinitely less likely.

Definition

A strategy profile σ is a **proper equilibrium** if it is the limit of a sequence σ^ε of ε-**proper equilibria**, where σ^ε is an ε-proper equilibrium if it is completely mixed and if $u_i\left(s_i, \sigma^\varepsilon_{-i}\right) < u_i\left(s_i', \sigma^\varepsilon_{-i}\right)$ then $\sigma_i^\varepsilon\left(s_i\right) < \varepsilon\sigma_i^\varepsilon\left(s_i'\right)$.

Colorful interpretation: players try harder to avoid worse mistakes. However, this should only lead to lower probabilities of worse mistakes, not infinitely lower (see Van Damme and Weibull 2002 for games with "control costs").

Instead, the point is that proper eqm captures backward induction without appealing to the agent-normal form, because deviating a second time incurs an addition cost and hence is much less likely.

Proper Equilibrium, Backward Induction, Iterated Weak Dominance

Proper equilibrium in the **normal** form implies perfect equilibrium in the **agent-normal** form (and hence implies sequential equilibrium).

Also implies iterated weak dominance:

	L	M	R
U	$1, 1$	$0, 0$	$-9, -9$
M	$0, 0$	$0, 0$	$-7, -7$
D	$-9, -9$	$-7, -7$	$-7, -7$

- (D, R) is weakly dominated, so not perfect.
- (M, M) is perfect: BR to equal trembles.
- But (M, M) isn't proper: if 2 plays M with prob near 1 and trembles s.t. M is 1's best response, then U is 1's second-best response. And if $\Pr(U) \gg \Pr(D)$ then 2 should take L, not M.

Proper Equilibrium: Existence

Theorem (Myerson 1978)

Every finite game has a proper equilibrium.

Proof. Similar to existence of perfect equilibrium. First, use Kakutani to show existence of ε-proper equilibrium for sufficiently small ε. Then take a convergence subsequence. More precisely:

- Let $m = \max_i |S_i|$ and let $\tilde{\Sigma}_i(\varepsilon) = \{\sigma_i : \sigma_i(s_i) \geq \varepsilon^m / m \ \forall s_i\}$.
- For small ε, $\tilde{\Sigma}_i(\varepsilon)$ is non-empty, compact, and convex.
- Let $r_i(\sigma) =$ $\left\{\sigma_i \in \tilde{\Sigma}_i(\varepsilon) : [u_i(s_i, \sigma_{-i}) < u_i(s_i', \sigma_{-i}) \implies \sigma_i(s_i) < \varepsilon \sigma_i(s_i')]\right.$.
- $r_i(\sigma)$ is non-empty, compact, convex, and has a closed graph.
- Apply Kakutani to get existence of ε-proper; take convergent subsequence.

Proper Equilibrium: Properties

Every proper equilibrium is perfect (and also perfect in the agent normal form).

Every strict equilibrium is proper.

Relation to sequential equilibrium:

Theorem (Kohlberg Mertens 1986)

A proper equilibrium of a normal-form game is a sequential equilibrium strategy profile in any extensive-form game with this (reduced) normal form.

- Reduced strategic form: delete redundant pure strategies.
- In this sense, proper equilibrium is invariant to extensive-form representation.
- Proper equilibrium refines sequential equilibrium without appealing to the agent-normal form, by implicitly making multiple deviations infinitely less likely.

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- Perfect equilibrium
- Proper equilibrium
- **Forward induction**
- Strategic stability
- Robustness of refinements

Forward Induction: Overview

The idea of forward induction is to interpret past off-path actions as deliberate attempts to influence future play.

Several alternative definitions. Illustrate with examples.

Forward Induction: Example

1 chooses T or B; if B, then U or D.
If 1 chooses B, 2 chooses L or R.

	L	R
T	$2,5$	$2,5$
BU	$4,1$	$0,0$
BD	$0,0$	$1,4$

- (T,R) and (BU,L) are both SPE.
- But BD is dominated by T, so forward induction says that if 2 sees 1 play B, 2 should infer that 1 will play U, and thus play L.
- If 2 reasons this way, 1 can always get 4 by playing BU. Thus, only (BU,L) is consistent with forward induction.

Burning Money and Iterated Weak Dominance

Another example (Ben-Porath Dekel 1992):

Stage 1: player 1 can either "Burn" or "Not Burn" 2.5 utils.

Stage 2: players 1 and 2 simultaneously play

	L	R
U	$9, 6$	$0, 4$
D	$4, 0$	$6, 9$

- $(Burn, D)$ is strictly dominated by $(NotBurn, D)$ (as $|u_1(D, L) - u_1(D, R)| < 2.5$).
- $(Burn, U)$ is not strictly dominated. (It's a BR to *(L if Burn, R is NotBurn)*.)
- Forward induction says that if 1 Burns, she "must" plan to play U. So 2 "should" play L after after Burn.
- Formally, any strategy where 2 takes R after Burn is deleted by **iterated weak dominance (IWD)**, as such strategies are weakly dominated if 1 never takes $(Burn, D)$. (**Not** strictly dominated.)

Burning Money (cntd.)

- Iterate again: 1 knows 2 will play L if 1 burns, so $(Burn, U)$ iteratively dominates $(NotBurn, D)$.
- Iterate again: since $(NotBurn, D)$ is iteratively dominated, if 2 sees $NotBurn$, 2 expects U, so 2 plays L.
- Iterate again: 1 plays $(NotBurn, U)$, 2 plays L.
- Punchline: 1 gets her best outcome without actually burning!

Slick argument, but IWD depends on order of deletion and has other robustness issues. . .

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- Perfect equilibrium
- Proper equilibrium
- Forward induction
- **Strategic stability**
- Robustness of refinements

Strategic Stability: Overview

Sequential/perfect/proper equilibria don't capture forward induction and (as we'll see) depend on arguably irrelevant details of the extensive form.

Motivated by these gripes, Kohlberg and Mertens (1986) define a notion of a stable set of NE.

- Set-valued in general, because set will be required to be robust to a large set of perturbations, and different perturbations will select different equilibria from the set.
- Recall: truly perfect equilibrium may not exist.

There are several versions of their definition and they are not directly applied much these days, but important for:

- subsequent refinements literature,
- inspiring signaling game refinements we'll cover next week,
- technical results on how NE set varies with payoff parameters,

so we briefly mention some main ideas.

Invariance (or Lack Thereof)

	a	b
a	$3,3$	$0,0$
b	$0,0$	$1,1$
Out	$2,2$	$2,2$

	a	b
a	$3,3$	$0,0$
b	$0,0$	$1,1$
Out	$2,2$	$2,2$
mix	$\frac{7}{3},\frac{7}{3}$	$\frac{4}{3},\frac{4}{3}$

- The games are "the same," with $mix = \frac{2}{3}Out + \frac{1}{3}a$.
- In the left matrix (Out, b) is proper, justified by 1 trembling to b more than a.
  $((a,a)$ is also proper.)
- In the right matrix, mix strictly dominates b. So 1 must tremble to mix much more than b. So (Out, b) is not proper. (Only (a,a) is proper.)

This shows that while proper equilibrium is invariant to adding or removing redundant pure strategies, it is not invariant to adding redundant mixed strategies "as pure strategies."

Stability Under All Perturbations

KM say that a set of NE is stable if, for any "perturbation of the game," the set contains a NE that is close to a NE of the perturbed game.

They consider different notions of stability. The most restrictive definition (which implies that sometimes only large sets can be stable, i.e., stability makes less precise predictiosn) requires stability to all payoff perturbations.

Definition

A set $\hat{\Sigma}(u)$ of NE of a game (I, S, u) is **stable under all perturbations** if for every $\varepsilon > 0$, there exists $\delta > 0$ s.t. for every payoff function u' with $\|u - u'\| < \delta$, there exists $\sigma \in \hat{\Sigma}(u)$ and a NE σ' of the game (I, S, u') s.t. $\|\sigma - \sigma'\| < \varepsilon$.

Stability Under All Perturbations (cntd.)

There is not always a single strategy profile with this property (because a truly stable equilibrium may not exist).

- If there is a strict NE, it satisfies the definition as a singleton.

The set of all NE satisfies the definition, by UHC. So there are "many" stable sets.

KM then say a set of NE is **hyperstable** if it stable under all perturbations but no proper subset of it is.

Stability

KM's main stability notion requires stability only for payoff perturbations that result from perturbing strategies (so smaller sets can be stable, i.e., stability can make more precise predictions than hyperstability).

Definition

A set $\hat{\Sigma}(u)$ of NE of a game (I, S, u) is **stable** if

1. for every $\varepsilon > 0$ and every completely mixed strategy profile σ^*, there exists $\bar{\delta} > 0$ s.t. for all $\delta = (\delta_1, \ldots, \delta_n) \in \left(0, \bar{\delta}\right)^n$ there exist $\sigma \in \hat{\Sigma}(u)$ and a NE σ' in the game where each player i is constrained to take σ_i^* with probability at least δ_i s.t. $\|\sigma - \sigma'\| < \varepsilon$; and
2. no proper subset of $\hat{\Sigma}(u)$ has this property.

If every element in a stable set yields the same (mixed) outcome, the outcome is **stable**.

Stability: Properties

All of KM's stability definitions depend on payoff perturbations in the normal form (not agent normal form) and the set of NE depends only on the reduced normal form, so the desired invariance property is built in.

By construction, stable sets do not contain weakly dominated strategies.

However, a stable set may fail to contain a sequential euqilibrium. (Hillas 1990 gives alternative definition without this problem.)

KM show that every stable set contains:

1. A stable set of any game obtained by deleting a weakly dominated strategy.
2. A stable set of any game obtained by deleting a strategy that is not a weak best reply to any opposing strategies in the set.

- The latter "NWBR" property can also be used as a stand-alone refinement, as we'll see when covering signaling games next week.

Outline

- Extensive-form games and subgame-perfect equilibrium
- Perfect Bayesian equilibrium
- Sequential equilibrium
- Perfect equilibrium
- Proper equilibrium
- Forward induction
- Strategic stability
- **Robustness of refinements**

Robustness of Equilibrium Refinements

Fudenberg Kreps Levine 1988 ask what happens if we interpret trembles as resulting from small incomplete information about players' preferences.

This yields a concept called *c*-**perfection**, which is similar to normal form perfection.

- Normal form—not agent normal form—because a deviation indicates that a player has an ex ante unlikely type that can predict later deviations as well.
- Thus, stronger refinements like sequential equilibrium or proper equilibrium implicitly assume that early deviations are mostly likely to be "random trembles," rather than indications that players have ex ante unlikely preference types.

Application to Implementation in Mechanism Design

Similar robustness critique applies to Nash vs. subgame-perfect implementation in mechanism design.

For a social choice function (SCF) is be fully Nash implementable (i.e., implementable in all NE) it must be (Makin) monotonic. However, more SCFs are implementable under refinements like SPE, because these refinements can eliminate undesirable equilibria (Moore Repullo 1988).

But this approach to eliminating undersirable equilibria is not robust to small incomplete information in FKL's sense.

Chung Ely 2003 show that only monotonic SCFs can be robustly implemented in undominated NE.

Aghion Fudenberg Holden Kunimoto Tercieux 2012 show that only monotonic SCFs can be robustly implemented in sequential equilibrium.

Robustness to All Incomplete Information Perturbations

FKL get a concept similar to normal form perfection by requiring robustness to **some** incomplete information perturbation

Kajii Morris (1997) require robustness to **all** incomplete information perturbations.

Not all games have robust equilibria in this sense (even if the game has a strict NE!).

However, if a game has a unique correlated equilibrium, it is robust.

For $p < 1/N$, a p-**dominant** equilibrium a is also robust, where p-dominance means that each a_i is a best reply to any conjecture μ_{-i} that puts probability at least p on a_{-i}.

Muhamet will cover robustness to incomplete information in detail in the second half of the course.