# Stable Observable Behavior

*Yuval Heller and Erik Mohlin (draft, September 5, 2014)*[*]

**Department of Economics, University of Oxford.**

**Abstract**

We study stable behavior in environments in which players are randomly matched to play a game, and before the game begins each player may observe how his partner behaved in the past. We show that strict Nash equilibria are always stable in such environments. We apply the model to study stable behavior in the Prisoner's Dilemma. We show that if players only observe past actions, then defection is the unique stable outcome. However, if players are able to observe past action profiles, then cooperation is also stable. Finally, we present an extension that studies the evolution of preferences.

**JEL Classification:** C72, C73, D01, D83.

**Keywords:** Random matching, indirect reciprocity, image scoring, secret handshake.

## 1 Introduction

In many economic situations people interact with other partners in short-term interactions. The lack of future interactions between the agents limit the ability to directly punish partners who acted unfairly, and incompleteness of contracts, non-verifiable information, court costs and other factors restrict the effectiveness of external enforcement. Agents in such interactions may obtain some information about the partner's behavior in a few past interactions with other partners, and may base their actions on this information. As a result, the behavior of agents may be influenced by the possibility of being observed by future partners and reputation considerations.

A few examples for such interactions include booking a holiday rental, making a one-time order from a remote trader, hiring a nanny, and visiting a tourist attraction. Information about

[*]Email addresses: yuval.heller@economics.ox.ac.uk and erik.mohlin@economics.ox.ac.uk. We would like to express our deep gratitude to Vince Crawford and Eddie Dekel for many useful comments.

the past behavior of the partner is often passed by word of mouth. Recently, such information is also conveyed by online sites that provide feedback from past interactions (e.g., eBay and Airbnb).

We study populations involved in such interactions, and characterize stable outcomes of a dynamic process of cultural learning. It is well known (see, e.g., Weibull, 1995) that without observability of past actions, stable outcomes are essentially Nash equilibria. *Our first main result shows that strict equilibria remain stable also when past behavior is observable. We then focus on the Prisoner's Dilemma, and show that defection is the unique stable outcome when agents only observe past actions, but that cooperation can be stable if they observe action profiles.*

**Basic Model.** Agents in a large population are randomly matched into pairs and play a symmetric one-shot game. Each agent before playing the game privately observes with probability $p$ an action played in the past by his partner, and observe a non-informative signal otherwise. We also consider noisy environments in which each agent observes with probability $\delta$ a "noisy" action sampled from an arbitrary distribution with full support (independent of the partner's past behavior.) The behavior of each agent is characterized by a *policy*: a mapping that assigns a mixed action for each possible message. The aggregate behavior of the population is described by a *strategy* $\sigma$*:* a distribution (with a finite support) over the set of policies. We interpret $\sigma$ as a population state in which $|C(\sigma)|$ policies coexist, each agent is endowed with one of these policies, and he follows this policy through out his life.

Our preliminary results (Section 2.4) show that each strategy induces a unique "consistent" outcome: a mapping that describes the distribution of actions played by each fraction of the population when being matched with each other fraction. This outcome determines the payoff of each policy in the population. We say that strategy $\sigma^*$ is evolutionary (neutrally) stable (Maynard Smith & Price, 1973) if for a sufficiently small $\epsilon$, incumbents (with mass $1-\epsilon$ ) who follow $\sigma^*$ strictly (weakly) outperform any group of agents ("mutants") with mass $\epsilon$ who follows any other strategy. An evolutionarily stable strategy, if adopted by a population in a given environment, cannot be invaded by any alternative strategy that is initially rare.

**Summary of Results.** An action $a^*$ is a (symmetric) strict equilibrium if $\pi(a^*, a^*) > \pi(a, a^*)$ for each action $a \neq a^*$. Our first result (Prop. 2) shows that any strict equilibrium of the underlying game is (1) neutrally stable, and (2) evolutionary stable in any noisy environment.[1] The intuition is that a homogenous population in which everyone plays $a^*$ (regardless of the observed signal) is stable: Any mutant who plays $a^*$ with probability lower than 1 is strictly outperformed if the

[1] In Section 3 we demonstrate that other refinements of Nash equilibrium are not necessarily stable. We show that the unique symmetric Nash equilibrium in the Hawk-Dove game (which is stable without observability, and satisfies all the standard refinements of Nash equilibrium) is not stable when $p > 0$.

mutants are sufficiently rare; mutants who differ in their behavior relative to the incumbents only after signals that are never observed (which is possible only in noiseless environments) play the same as the incumbents and obtain the same payoff.

We next consider *"separable" Prisoner's dilemma games* in which each player decides simultaneously whether he cooperates by giving up $g > 0$ of his own payoff, in order to yield $1 + g$ fitness points to his partner (see Table 2, right side). We show (Prop. 3) that any neutrally stable strategy must be pure (either everyone cooperating or everyone defecting), and that in noisy environments defection is the unique stable outcome. The intuition is that the opponent's probability of cooperation is a linear function of the player's probability of cooperation, which implies that the payoff of a policy is a linear function of its induced probability of cooperation. If this function is weakly decreasing (strictly increasing), then the incumbents must never (always) cooperate, or otherwise, mutants that never (always) cooperate would outperform the incumbents. When there is noise ($\delta > 0$), cooperation with probability one implies cooperation regardless of the signal, but then mutants who always defect yield a strictly higher payoff.

**General Model.** Our next step is to extend the previous results to a general observability structure in which players may observe several past actions and past action-profiles, and to a general noise structure which includes implementation errors (and not only observation errors.) This generalization raises a few technical difficulties, which are dealt with in Sections 7-8.[2] When players observe several past actions, a strategy may admit several consistent outcomes. Thus, we define a configuration as a pair consisting of a strategy and a consistent outcome, and adapt the notion of evolutionary stability to configurations. A configuration is *strongly (weakly) evolutionary stable* if for sufficiently small $\epsilon$, after an invasion of $\epsilon$ mutants who play a different strategy in every (there exists an) outcome that is consistent with the post-entry state: (1) the incumbents play the same among themselves (i.e., rare mutants do not substantially change the aggregate behavior; a focality condition a la Dekel *et al.* (2007)), and (2) the mutants are outperformed. We extend the noise structure by considering perturbed games in which players may rarely tremble when they choose policies or actions (a la Selten (1975, 1983).) A strategy is a *limit strong (weak) evolutionary stable* if it is the limit of strong (weak) evolutionary stable strategies in a converging sequence of perturbed game.

We extend both of the main results mentioned above to the general model (Theorems 1-2), and we show that: (1) any strict equilibrium is limit strong evolutionary stable given any observability structure,[3] and (2) defection is the unique limit weak evolutionary stable outcome in separable

[2] Due to these technical difficulties, the general model may be somewhat harder to read, and w postpone its formal presentation to the last part of the paper.

[3] Moreover it is robust in the sense that it is the limit of strong evolutionary stable configurations in *any*

Prisoner's Dilemma game if players observe actions (rather than action profiles.)

Our next result shows that cooperation can be sustained as a stable outcome in the Prisoner's Dilemma if players observe past action profiles. Specifically, we show that there is a limit strong evolutionarily stable configuration in which players always cooperate. The intuition is that players defect after observing that the opponent was the sole defector in the past (i.e., they observe an action profile of $(d, c)$), and cooperate otherwise. The stability holds for perturbations in which players rarely "tremble" by choosing the policy that always defects.

**Most related Literature** Robson (1990) presented the mechanism of "secret handshake" that can take a population from any inefficient state (say $\sigma^*$) to an efficient state (say $\sigma'$). According to this mechanism, a small group of mutants sends a special signal before interacting; the incumbents are assumed to ignore this signal and always play $\sigma^*$, while the mutants usually play $\sigma^*$, unless both players have sent the special signal, and in this case they both play $\sigma'$. Several papers applied this mechanism, and showed that with "cheap talk" with sufficiently rich language (e.g., Kim & Sobel (1995)) or when players can observe each other's subjective preferences with sufficiently high probability (e.g., Dekel *et al.* (2007)), then stability implies efficiency. Our first main result (Theorem 1) shows that the "secret handshake" mechanism is not effective when players observe past behavior: non-efficient strict equilibria are stable. As argued in Section 6, this casts doubts on the use of "secret handshake" arguments in the literature on the evolution of preferences.

In an influential paper, Nowak & Sigmund (1998), present the mechanism of *image scoring* to sustain cooperation in the Prisoner's Dilemma through indirect reciprocity (Trivers (1971)). In this mechanism each player observes several past actions of the partner, and he cooperates iff the partner's frequency of cooperation is above some threshold (see also the recent extension in Berger & Grüne (2014)). Our second main result (Theorem 2) generalizes and formalizes existing criticism of "image scoring" (Leimar & Hammerstein (2001); Panchanathan & Boyd (2003)) and shows that no mechanism that relies only on observing past actions can sustain cooperation (unless the set of feasible strategies is severely restricted, see Remark XX).

In a seminal work Sugden (1986) presented the mechanism of *good standing* (see also its generalization in Kandori (1992)). In this mechanism, all agents are initially in good standing; an agent looses good standing by defecting against a partner in good standing (whereas defecting against a partner in bad standing does not damage the player's standing); an individual lacking good standing can regain it by cooperating. In this mechanism players have to asses the standing of the partner, and this assessment is very demanding both informationally and cognitively, as it

converging sequence of perturbations in which players tremble when choosing actions (i.e., it is *strict* limit strong evolutionary stable.)

depends on long histories of play of many players (because the reputation of one agent depends on the reputations of his past partners.) Our third main result (Theorem 3) presents a novel mechanism to support cooperation, in which each player only has to observe a single past action profile of his partner, without requiring any higher order information.

**Extensions.** We present three extensions of our model. In Section 10, we consider an alternative setup in which both signals are observed by both players. Such a setup may fit better feedback mechanisms in on-line sites (e.g., eBay) in which the feedback on the past behavior of the agents is public. Minor adaptations to the proofs show that all the results of the main model hold also in this setup. In addition, we show (Theorem 4) that public signals allow to support cooperation as a stable outcome in a more robust way that holds regardless of the source of noise. In Section 5 we endogenize the observation probability. Specifically, we assume that each player chooses an effort level that determines his probability to observe a past action of the partner. All of our results also hold in this extension.[4]

Finally, we enrich the model to study the evolution of preferences (Section 6). Specifically, we assume that each agent is endowed with subjective preferences that may differ from the objective fitness. A population state is a a pair: a distribution of preference types, and a policy for each type. A state is consistent if all agents use policies that maximize their subjective preferences. A state is stable, if any sufficiently small group of mutant types is outperformed in any "focal" post-entry state (in which the incumbents maintain their pre-entry policies). It is possible to extend all our results to this setup, and characterize which subjective preferences are stable. Finally, we relate our approach with the literature on the evolution of preferences, and argue why our approach suggests a novel solution to inherent problems in that literature.

**Structure.** Section 2 presents the basic model (observing a single action). In Section 3 we prove that strict equilibria are stable. Section 4 shows that defection is the unique stable outcome in the Prisoner's Dilemma. Next, we present two extensions: endogenous observability (Section 5), and subjective preferences (Section 6.) Section 7 presents our general model in which players may observe several actions and action profiles. In Section 8 we develop solution concepts that deal with multiple consistent outcomes and with general forms of noise. Section 9 presents the results of the general model. In Section 10 we study public signals. Section 11 surveys the related literature. We conclude in Section 12.

[4] Moreover, we show (Section 5.2) that under the additional assumption that actions are unobservable without spending an effort , then there are strong connections between stability and Nash equilibria of the underlying games: (1) any pure stable outcome must be a Nash equilibrium, and (2) any evolutionary stable strategy of the underlying game (without observability) remains stable in this setup.

# 2 Basic Model (Observing a Single Action)

In this section we present our basic model, in which each player may privately observe a past action of the opponent. The second part of the paper (starting in Section 7) extends the model to deal with observation of several actions and action profiles.

## 2.1 Underlying Game

We present a reduced form static analysis of a dynamic evolutionary process of cultural learning in a large population of agents.[5] The agents in the population are randomly matched into pairs and play a symmetric one-shot game $G$.[6] Formally, let $G = (A, \pi)$ be a two-player symmetric normal-form game, where $A$ is a finite set of actions for each player, and $\pi : A \times A \to \mathbb{R}$ is the payoff function. We interpret, as is standard in the evolutionary game theory literature, the payoffs as representing "success" or "fitness". Let $\Delta(A)$ denote the set of mixed actions (distributions over $A$), and let $\pi$ be extended to mixed actions in the usual way. We use the letter $a$ ($\alpha$) to denote a typical pure (mixed) action. With slight abuse of notation let $a \in A$ also denote the element in $\Delta(A)$, which assigns probability 1 to $a$, and we adopt this convention for all probability distributions throughout the paper.

## 2.2 Observations and Strategies

An *observation structure* (or environment) is a triple $(p, \delta, \tilde{\alpha})$ where: (1) $0 \leq p \leq 1$ is the probability of observing a past action of the opponent (or partner) before playing the game; (2) $0 \leq \delta \leq 1 - p$ is the probability of observing a "noisy" action; and (3) $\tilde{\alpha} \in \Delta(A)$ is a distribution with a full support (i.e., $\tilde{\alpha}(a) > 0$ for each $a \in A$) from which noisy actions are sampled. That is, before playing the game, each player privately observes: (1) with probability $p$ an action played in the past by his opponent, (2) with probability $\delta$ an action sampled from the noise distribution $\tilde{\alpha}$, and (3) with the remaining probability $(1 - p - \delta)$ a non-informative signal $\phi$. We say that an observation structure is *noisy* if $\delta > 0$ and *noiseless* if $\delta = 0$. We say that an observation structure is *perfect* if $p = 1$ and *imperfect* if $p < 1$.

Let $M = A \cup \phi$ be the set of possible messages (or signals), and let $m$ denote a typical message (element of $M$). A *policy* $s : M \to \triangle(A)$, is a mapping that assigns a mixed action for each

[5] The formal model also captures a biological evolutionary process in which the behavior of the agents is influenced by genetic inheritance.

[6] The assumption that the underlying game $G$ is symmetric is essentially without loss of generality. As standard in the evolutionary literature (e.g., Selten, 1980), asymmetric contests can be symmetrized by looking at a symmetric game in which an agent chooses an action at the ex ante stage (before being assigned to one of the roles in the game).

possible message. We interpret $s_m(a) = s(m)(a)$ as the probability that a player who follows policy $s$ plays action $a$ after observing message $m$. We use the notation $a$ to denote also the policy that always plays action $a$ regardless of the signal.

Let $S$ denote the set of all policies, and let $\Sigma \equiv \triangle(S)$ denote the set of distributions over the set of policies with a finite support. A *strategy* (or *population state*) $\sigma \in \Sigma$ is a distribution (with a finite support) over the set of policies. Let $\sigma(s)$ denote the probability that strategy $\sigma$ assigns to policy $s$. Given a strategy $\sigma \in \Sigma$, let $C(\sigma)$ denote its support (or carrier, i.e., the set of policies such that $\sigma(s) > 0$.) We interpret such population state $\sigma$ as describing a population in which $|C(\sigma)|$ policies coexist, each agent is endowed with one of these policies according to the distribution of $\sigma$, and he follows this policy through out his life. Given this interpretation, we will identify elements in $\Sigma$ as population states. When $|C(\sigma)| = 1$, we identify the strategy with the unique policy in its support.

Given message $m \in M$, let $\sigma_m \in \Delta(A)$ denote the distribution of actions that a random policy sampled from $\sigma$ plays after observing message $m$:

$$\sigma_m(a) = \sum_{s \in C(\sigma)} \sigma(s) \cdot s_m(a).$$

## 2.3 Outcomes

Given a finite set of policies $\tilde{S} \subset S$, an *outcome* $\eta : \tilde{S} \times \tilde{S} \to \triangle(A)$ is a mapping that assigns to each pair of policies $s, s' \in \tilde{S}$ a mixed action $\eta_s(s')$, which is interpreted as the mixed action played by a player with policy $s$ conditional on being matched with an opponent with policy $s'$. Let $O_{\tilde{S}} \equiv (\triangle(A))^{\left(\tilde{S} \times \tilde{S}\right)}$ denote the set of all outcomes defined over the set of policies $\tilde{S}$.

We present a few definitions for a given strategy $\sigma \in \Sigma$, an outcome $\eta \in O_{C(\sigma)}$, and a policy $s \in C(\sigma)$. Let $\eta_{s,\sigma} \in \triangle(A)$ be the mixed action played by a player with policy $s$ when being matched with a random opponent sampled from $\sigma$. Formally, for each action $a \in A$:

$$\eta_{s,\sigma}(a) = \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s')(a).$$

Let $\eta_\sigma \in \triangle(A)$ be the aggregate mixed action player by a random player (sampled from $\sigma$) being matched with a random opponent. Formally, for each action $a \in A$:

$$\eta_\sigma(a) = \sum_{s \in C(\sigma)} \sigma(s) \cdot \eta_{s,\sigma}(a).$$

Let $\mu_{s,\sigma,\eta} \in \triangle(A \times A)$ be the (possibly correlated) mixed action profile that is played when

a player with policy $s$ is matched with a random opponent sampled from $\sigma$. Formally, for each action profile $a, a' \in A \times A$ (where $a$ is interpreted as the action of the player with policy $s$, and $a'$ as the action of his opponent):

$$\mu_{s,\sigma,\eta}(a, a') = \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s')(a) \cdot \eta_{s'}(s)(a').$$

## 2.4 Consistent Outcomes

Fix an observation structure $(p, \delta, \tilde{\alpha})$. Given a strategy $\sigma \in \Sigma$, let $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ be the mapping between outcomes that is induced by $\sigma$ (and the observation structure.) That is, the outcome $f_\sigma(\eta)$ is interpreted as the new outcome induced by players with policies that are sampled from $\sigma$, who are randomly matched, and, before playing, observe a signal according to the observation structure with respect to the past behavior of the opponent that is induced by the outcome $\eta$. Formally:

$$\begin{aligned}(f_\sigma(\eta))_s(s')(a) &= (1-p-\delta) \cdot s(\phi)(a) + p \cdot \sum_{a' \in A} \eta_{s',\sigma}(a') \cdot s(a')(a) \\ &\quad + \delta \cdot \sum_{a' \in A} \tilde{\alpha}(a') \cdot s(a')(a)\end{aligned}$$

An outcome $\eta \in O_{C(\sigma)}$ is consistent with strategy $\sigma$ if it is a fixed point of this mapping: $f_\sigma(\eta) \equiv \eta$. The following standard lemma shows that each strategy admits consistent outcomes.

**Lemma 1.** *For each strategy $\sigma \in \Sigma$ there exists a consistent outcome $\eta$.*

*Proof.* Observe that the space $O_{C(\sigma)}$ is a convex and compact subset of an Euclidean space, and that the mapping $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ is continuous. Brouwer's fixed-point theorem implies that the mapping $\sigma$ has a fixed point $\eta*$, which is a consistent outcome by definition.

The following simple example shows that with perfect observability ($p = 1$) a strategy may admit many consistent outcomes. □

**Example.** Assume that the underlying game has two actions $A = \{c, d\}$. Assume perfect observability (i.e., $p = 1$ ). Let $\tilde{s}$ be the following "tit-for-tat" policy: $\tilde{s}(a) = a$ for each $a \in A$ (i.e., each player plays the observed past action of his opponent). Observe that any outcome $\eta \in O_{\tilde{s}}$ (i.e., a probability to play $c$) is consistent with the strategy $\tilde{s}$ (which assigns mass 1 to policy $\tilde{s}$.) Note that if $p < 1$ then it s relatively simple to show that there exists a unique consistent

outcome, which is given by:

$$\eta = \frac{\delta}{1-p} \cdot \tilde{\alpha} + \frac{1-p-\delta}{1-p} \cdot \tilde{s}(\phi).$$

Next we show that with imperfect observability, consistent outcomes are unique.

**Proposition 1.** *Assume $p < 1$. Let $\sigma \in \Sigma$ be a strategy. If $\eta, \eta' \in O_{C(\sigma)}$ are consistent outcomes, then $\eta = \eta'$.*

The intuition is as follows (the proof is in Appendix A.1.) Consider two cases: in which a player with a random policy sampled from strategy $\sigma$, observes a past action of the opponent sampled according to $\eta$ in case I and sampled according to $\eta'$ in Case II, and then plays an action according to his sampled policy. The played action differs in these two cases only if (1) an opponent past action is observed (which occurs with probability $p < 1$), and (2) the observed actions are different (which is, roughly speaking, the "distance" between the outcomes $\eta$ and $\eta'$). This implies that the distance between $f_\sigma(\eta)$ and $f_\sigma(\eta')$ is strictly smaller than the distance between $\eta$ and $\eta'$ (i.e., $f_\sigma$ is a contraction mapping). Thus, two different outcomes cannot be both consistent with the same strategy.

Observe that the above argument relies on the observation of a single action. If several actions are observed, then the probability that the observed action profiles differ may be larger than the distance between the two outcomes (as demonstrated in the example in Section 7). In order to avoid the added complexity of having multiple consistent outcomes, we assume throughout the basic model that only a single action is observed, and that the observation probability $p$ is smaller than one. This implies that each strategy $\sigma$ induces a unique consistent outcome, which we denote by $\eta(\sigma)$.

## 2.5 Payoffs

Given a strategy $\sigma \in \Sigma$ and a policy $s \in C(\sigma)$ let $\pi_s(\sigma)$ be the payoff of a player who follows policy $s$ in population state $\sigma$. Formally:

$$\pi_s(\sigma) = \sum_{(a,a') \in A \times A} \pi(a, a') \cdot \mu_{s,\sigma,\eta(\sigma)}(a, a').$$

Given a strategy $\sigma'$ with a smaller support than $\sigma$ ($C(\sigma') \subseteq C(\sigma)$), let $\pi_{\sigma'}(\sigma)$ be the payoff of a player with a policy sampled according to $\sigma'$ in population state $\sigma$:

$$\pi_{\sigma'}(\sigma) = \sum_{s' \in C(\sigma')} \sigma'(s') \cdot \pi_{s'}(\sigma).$$

We say that a strategy is *balanced* if $\pi_s(\sigma) = \pi_{s'}(\sigma)$ for every $s, s' \in C(\sigma)$, in that case, we write the uniform payoff as $\pi(\sigma)$.

## 2.6 Evolutionary Stability

Our solution concept is Maynard Smith & Price (1973)'s notion of stability, which requires resistance to small groups of agents who experiment with new behavior.

Given two strategies $\sigma^*, \sigma' \in \Sigma$ with relative masses of $(1-\epsilon)$ and $\epsilon$ let $\sigma_{\sigma^*,\epsilon,\sigma'}$denote the *mixture strategy* (or *post-entry state*). Formally, for each $s \in C(\sigma) \cup C(\sigma')$:

$$\sigma_{\sigma^*,\epsilon,\sigma'}(s) = (1-\epsilon) \cdot \sigma^*(s) + \epsilon \cdot \sigma'(s).$$

A strategy is evolutionary (neutrally) stable if after any invasion of a small group of players (mutants) who play a different strategy, the incumbents (who follow the original strategy) achieves a strictly (weakly) higher payoff than the mutants. Formally:[7]

**Definition 1.** (Maynard Smith & Price (1973); Maynard-Smith (1982)) Strategy $\sigma^*$ is evolutionary (neutrally) stable if there exists $\bar{\epsilon} > 0$ such that for each strategy $\sigma'$ and each $0 < \epsilon < \bar{\epsilon}$ the following inequality holds:

$$\sigma' \neq \sigma^* \Rightarrow \pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}) < \pi_{\sigma^*}(\sigma_{\sigma^*,\epsilon,\sigma'}) \quad (\pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}) \leq \pi_{\sigma}(\sigma_{\sigma^*,\epsilon,\sigma'})).$$

The motivation for the above definition is that an evolutionarily stable strategy, if adopted by a population of players, cannot be invaded by any alternative strategy that is initially rare. It is well known that evolutionary stable strategies are dynamically stable in smooth payoff-monotonic selection dynamics (e.g., Taylor & Jonker (1978); Cressman (1997); Sandholm (2010)), and that any neutrally stable strategy is balanced (see also Claim 3.)

# 3 Stability of Strict Equilibria

In this section we show that strict equilibria are evolutionary stable for any noisy observation structure (and neutrally stable for a noiseless observation structure.)

The intuition is as follows. Consider a strategy $s^* \equiv a^*$ in which all the incumbents play $a^*$ regardless of the observed message. In any noisy observation structure all messages are observed with positive probability. As a result, any mutant strategy who plays $a^*$ with probability lower

[7] In order to be consistent with latter definitions (in particular, with the definition of uniform limit evolutionary stable strategy), we use the reformulation of the original definitions with a uniform barrier (see, e.g., (Weibull, 1995, Proposition 2.5).)

than 1 after any observed message, is strictly outperformed against the incumbents, and thus is strictly outperformed if the mutants are sufficiently rare. This implies that $a^*$ is evolutionary stable. In noiseless observation structures, mutant strategies that differ from $a^*$ only after messages that are never observed (i.e., messages different than $a^*$ and $\phi$), induce the same play as the incumbents, and thus in this case, $a^*$ is only neutrally stable. The formal statement of the result is as follows:

**Definition 2.** Action $a^* \in A$ is a strict equilibrium of the underlying game $G = (A, \pi)$ if for each action $a \neq a^*$: $\pi(a^*, a^*) > \pi(a, a^*)$.

**Proposition 2.** *If action $a^*$ is a strict equilibrium then the strategy $a^*$ is: (1) neutrally stable, and (2) evolutionary stable if $\delta > 0$.*

The proof is omitted because Prop. 2 is a special case of Theorem 1, which shows that a strict equilibrium is evolutionary stable also if players may observe several past actions.

An interesting question is whether Prop. 2 can be strengthened to weaker solution concepts than strict equilibrium. The following example shows that this is not the case: the unique symmetric Nash equilibrium of the underlying game is not stable in any environment with $p > 0$. Note that a unique symmetric equilibrium of a symmetric game satisfies all the standard (non-evolutionary) refinements of Nash equilibrium.

Tab. 1: An Example of a Hawk-Dove (Chicken) Game

| | $d$ | $h$ |
|---|---|---|
| $d$ | 1, 1 | 0.5, 1.5 |
| $h$ | 1.5, 0.5 | 0, 0 |

**Example.** Consider the game of Hawk-Dove (or Chicken) presented in Table 1 in which the players have two actions: $d$ ("dove") and $h$ (hawk), each action is the strict best-reply to the other action, and $\alpha^* = (0.5, 0.5)$[8] is the unique symmetric Nash equilibrium of the underlying game[9] (and it is an evolutionary stable strategy in the underlying game.) We now show why strategy $\alpha^*$ is not neutrally stable when past action are observed with any positive probability $p$. To simplify the argument we assume noiseless environment ($\delta = 0$.) Consider mutant strategy

[8] As standard in the literature, $(\alpha_1, ..., \alpha_n)$ denotes the mixed action that assigns probability $\alpha_i$ to the $i$-th action.

[9] Asymmetric equilibria cannot be played in our setup in which player they cannot condition their play on being the row/column player.

that assign equal weights to three policies: (1) always playing $h$, (2) always plying $d$, and (3) a policy that plays the opposite action to the observed message (and plays each action with equal probability if observed $\phi$). These mutants obtain the same payoff against the incumbents (because all actions yield the same payoff against $\alpha^*$), but a strictly higher payoff against other mutants (because when two mutants are matched they play the inefficient action profile $(h,h)$ with probability of only $\left(\frac{1}{3}\right)^2+\left(\frac{1}{3}\right)^2\cdot\frac{1}{4}<\frac{1}{4}$, while when an incumbent and a mutant are matched they play all action profiles, including the inefficient $(h,h)$, with probability $\frac{1}{4}$). The example can be extended to any Hawk-Dove game.

# 4 Prisoner's Dilemma

In this section we show that only mutual defection is stable in the Prisoner's Dilemma.

## 4.1 Underlying Game

Tab. 2: General Prisoner's Dilemma $(1+l>g,l>0\ )$ and Separable Prisoner's Dilemma

| | $c$ | $d$ |
|---|---|---|
| $c$ | $1$ $\quad 1$ | $-l$ $\quad 1+g$ |
| $d$ | $1+g$ $\quad -l$ | $0$ $\quad 0$ |

General Prisoner's Dilemma

| | $c$ | $d$ |
|---|---|---|
| $c$ | $1$ $\quad 1$ | $-g$ $\quad 1+g$ |
| $d$ | $1+g$ $\quad -g$ | $0$ $\quad 0$ |

Separable Prisoner's Dilemma

The left side of Table 2 presents a general Prisoner's Dilemma game that depends on two parameters $0<g,l$. When both players play action $c$ (*cooperate*) they both get a high payoff (normalized to one), and when they both play action $d$ (*defect*) they get a low payoff (normalized to zero). When a single player defects he obtains a payoff of $1+g$ (i.e., he gains an additional payoff of $g$) while his opponent gets $-l$ (a loss of $l$.) Note the $(c,c)$ is the unique efficient outcome that maximizes the sum of payoffs (because $g<1+l\Rightarrow 1+g-l<2$), while $d$ is strictly dominant and $(d,d)$ is the unique Nash equilibrium (and a strict equilibrium) of the game.

The right side of Table 2 presents a Prisoner's dilemma games in which the gain obtained by defecting is independent of the opponent's action (i.e., $g=l$). These games represent "separable" interactions in which each player decides simultaneously if he give up $g<1$ of his own payoff, in order to yield $1+g$ fitness point to his partner (see, e.g., Hamilton (1964) and Van Veelen (2009), for the discussing applications of these games and their importance in evolutionary models.)

## 4.2 Only Defection is Stable

In this subsection we restrict attention to separable Prisoner's Dilemma and show that essentially always defecting is the unique neutrally stable strategy (which is also evolutionary stable in any noisy environment due to Prop. 2.)[10] Our result includes two parts: (1) any strategy that induces a non-pure outcome cannot be neutrally stable, and (2) only defection (but not cooperation) is neutrally stable in any noisy environment.[11]

**Proposition 3.** *Assume that $G$ is a separable Prisoner's Dilemma game, and that $\sigma^*$ is a neutrally stable strategy with consistent outcome $\eta^*$. Then:*

1. $\eta^*_{\sigma^*}(d) \in \{0,1\}$ *(i.e.,the induced outcome is pure.)*
2. *If $\delta > 0$, then $\sigma^* \equiv d$.*

The sketch of the proof is as follows (the formal proof is in Appendix A.2.) In a separable Prisoner's Dilemma, a player obtains $1+g$ points if his opponent defects, and he suffers a cost of $g$ points if he cooperates. The probability that the opponent defects depends on the message the opponent observes. This implies that the payoff of a policy depends only on its probability of defection (which determines the probability that the opponent observes cooperation rather than defection), and it is a linear function of this probability. This implies that either:

1. The payoff is weakly increasing in the probability of defection, which implies that mutants who defect with probability one will outperform the incumbents (they fare weakly better against incumbents, and strictly better against other mutants;) or
2. The payoff is strictly decreasing in the probability of defection, which implies that the incumbents must cooperate with probability one (otherwise, mutants who always cooperate would outperform the incumbents.) When there is noise ($\delta > 0$), it implies that the incumbents must cooperate after any observed signal, but this leads to a contradiction (the payoff cannot be decreasing in the probability of defection, if everyone cooperates after any observed signal.)

[10] The proof relies on the separability of the payoffs which simplifies the analysis. In the next revision of this paper we will extend the result to the of $l \leq g$, as well as analysis the case of $g < l$ (in which we present a novel variant of "image scoring" to sustain cooperation.

[11] Prop. 3 is related to an existing result of Ohtsuki (2004, Section 4). Ohtsuki (2004) studies a setup in which players are randomly matched to play the Prisoner's Dilemma, each player observes his opponent's last played action with positive probability, and conditions his play only on this signal (but not on his own past behavior.) He shows that in the presence of positive level of noise defection is the unique stable outcome in a specific evolutionary dynamics (the "adaptive dynamics" in which the population continuously moves towards the direction of the local best-response to the current population state). Our result is stronger as it does not depend on a specific dynamics.

*Remark* 1. Several papers study equilibria in repeated Prisoner's Dilemma with private imperfect monitoring of the opponent's past actions, and they present "folk theorem" results in this setup when the signals are sufficiently accurate and the players are sufficiently patient (see the survey of Kandori, 2002). Most of these equilibria (e.g., Sekiguchi, 1997; Ely & Välimäki, 2002) rely on inducing the players to be indifferent between defection and cooperation after all histories. While the setup is different than ours, we note that simple adaptations of the above arguments can show that no such equilibrium is neutrally stable: an arbitrarily small group of "mutants" who always defect would strictly outperform incumbents who follow such an equilibrium.

# 5 Endogenous Observability

In this section we extend our analysis to a setup in which the observation probability is endogenously determined by the players. The results of this section are not used later in the paper (and readers not interested in the endogenous observability may skip to the next section.)

## 5.1 Changes to the Basic Model

The behavior of each player, is characterized by a *type* - a pair $(e, s)$, where $e \in \mathbb{N}$ is the effort level, and $s$ is the policy. The fixed observation probability $p$ in the basic model is replaced with a weakly increasing observability function $p : N \to [0, 1)$ with the following interpretation: a player who spends effort $e$ observes a past action of his opponent with probability $p(e)$. Let $T = \mathbb{N} \times S$ be the set of all types (pairs of effort and policy), with a typical element $t$.[12] Given a type $t$, let $e(t)$ ($s(t)$) denote its effort (policy.) We redefine a strategy to be a distribution (with a finite support) over $T$. We adapt the definition of an outcome and a consistent outcome in a straightforward way (replacing the constant $p$ with the function $p(e(t))$.) Minor adaptations to Lemma 1 and Prop. 1 show that each strategy admits a unique outcome also in this setup. Finally, the payoff of a type is adapted by reducing the cost of the effort from the payoff, i.e.,:

$$\pi_t(\sigma) = \sum_{(a,a') \in A \times A} \pi(a, a') \cdot \mu_{t,\sigma,\eta(\sigma)}(a, a') - c(e(t)),$$

where $c : \mathbb{N} \to \mathbb{R}^+$ is a strictly increasing cost function. We normalize the cost of zero effort to be zero: $c(0) = 0$. We adapt the definitions of neutral and evolutionary stability in a straightforward way. All of the results presented in the previous sections (Prop. 2-3) hold also in this extended setup with minor adaptations to the proofs.

[12] Our result remain qualitatively the same with a continuum of feasible efforts, except that the stability in Claim 2 is obtained without a uniform invasion barrier.

## 5.2 Additional Results when Observability Requires Effort

In this subsection we show that if one assumes that observability of past actions requires some effort (i.e., $p(0) = 0$), then there are strong connections between stability and Nash equilibria of the underlying game: (1) any pure stable outcome is a Nash equilibrium, and (2) any evolutionarily stable strategy of the underlying game remains stable in our setup.

### 5.2.1 Pure Stable Outcomes are Nash Equilibria

We now show that if a neutrally stable strategy induces a pure outcome, then this outcome must be a Nash equilibrium of the underlying game. Together with Prop.1, it implies that strategies which induce pure outcomes are stable essentially if and only if they are Nash equilibria (the "if" direction requires being strict Nash.)

We begin by presenting a simple lemma, that shows that in any neutrally stable strategy with a pure outcome no one spends efforts. The intuition is that otherwise a mutant strategy that plays the same pure outcome, but spends no effort, would strictly outperform the incumbents.

**Definition 3.** Let $\eta$ be an outcome consistent with strategy $\sigma$. The outcome $\eta$ is *pure* if a single action is played with probability 1: i.e., $\exists a^* \in A$ s.t. $\eta_\sigma(a^*) = 1$.

**Lemma 2.** *Let $\sigma^*$ be a neutrally stable strategy with a consistent pure outcome . Then $t \in C(\sigma)$ implies $e(t) = 0$.*

*Proof.* Assume to the contrary that there exists type $t^* \in C(\sigma^*)$ with $e(t^*) > 0$. Consider the mutant type $t' = (0, a^*)$ that spends no effort and always plays $a^*$. Observe that in the post-entry state, everyone plays $a^*$ and thus the payoff of each type is $\pi_t(\sigma) = \pi(a, a') - c(e(t))$. This implies that: $\pi_{t'}(\sigma) = \pi(a, a') > \pi_\sigma(\sigma)$, and we get a contradiction to the neutrally stability. □

Next, we show that if $p(0) = 0$, then a pure outcome of a neutrally stable strategy must be a Nash equilibrium. The intuition is that if the outcome is not a Nash equilibrium, then a mutant who best replies would strictly outperforms the incumbents (who continue to play the same action also against the mutant because they spend no effort and do not observe informative signals.)

*Claim* 1. Assume that $p(0) = 0$. Let $\sigma^*$ be a neutrally stable strategy with a consistent pure outcome $\eta^*$. Assume that $\eta^*_{\sigma^*}(a^*) = 1$. Then $a^*$ is a Nash equilibrium of the underlying game.

*Proof.* Assume to the contrary that $a^*$ is not a Nash equilibrium. Let $a' \in A$ be a better reply to $a^*$: $\pi(a', a^*) > \pi(a^*, a^*)$. Consider a mutant type $t = (0, a')$ that spends no effort and always plays $a'$. Lemma 2 implies that the incumbent types spend no efforts and do not observe informative massages. This implies that the incumbents also play $a^*$ against the mutant

in any post-entry state. Thus, the mutant type achieves a strictly higher payoff relative to the incumbents when facing an incumbent. This implies that if the mutants are sufficiently rare, they strictly outperform the incumbents in the post entry state. □

### 5.2.2 Stability of ESSs of the Underlying Games

We conclude this section by showing that any neutrally (evolutionarily) stable strategy of the underlying game ($\alpha^*$) is also a neutrally (essentially an evolutionarily) stable strategy in our environment. The intuition is as follows. The stability of $\alpha^*$ in the underlying game implies stability against mutants who spends no effort; if a mutant spend a positive effort, it can only gain more than the incumbents when facing a mutant, and if the mutants are sufficiently rare they are strictly outperformed due to the effort's cost. The only mutants which can obtain the same payoff as the incumbents in the post-entry state are those that are essentially equivalent to the incumbents: they play the same marginal distribution $\alpha^*$, but possibly condition their play on the observed noisy actions.[13] Formally:

*Claim* 2. Assume that $p(0) = 0$.

1. Let $\alpha^* \in \Delta(A)$ be a neutrally stable strategy of the underlying game. Then $\sigma^* = (0, \alpha^*)$ is a neutrally stable strategy of the environment with observability.

2. Assume that $\delta > 0$. Let $\alpha^* \in \Delta(A)$ be an evolutionarily stable strategy of the underlying game. Then there exists $\bar{\epsilon} < 1$, such that for each $0 < \epsilon < \bar{\epsilon}$ and $\sigma' \in \Sigma$ if $\pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}) = \pi_{\sigma^*}(\sigma_{\sigma^*,\epsilon,\sigma'})$, then $(\eta(\sigma_{\sigma^*,\epsilon,\sigma'}))_{\sigma'} = \alpha^*$.

*Proof.*

1. Consider an arbitrary mutant strategy. The fact that $\alpha^*$ is a neutrally stable strategy of the underlying game implies neutral stability against mutants who do not spend efforts. If the mutants spend a positive effort, and the mutants are sufficiently rare, then the effort's cost (which is at least $c(1) > 0$) outweighs the potential gain (which can only be gained when two mutants are matched together).

2. Consider any mutant strategy $\sigma'$. The fact that $\alpha^*$ is an evolutionarily stable strategy of the underlying game implies that if the mutants are sufficiently rare, then they obtain a strictly lower payoff than the incumbents, unless their marginal play is the same as the incumbents (i.e., $(\eta(\sigma_{\sigma^*,\epsilon,\sigma'}))_{\sigma'} = \alpha^*$).

□

[13] One can show that the set of strategies that equivalent to $(0, \alpha^*)$ in that sense is an evolutionarily stable set a la Thomas (1985).

# 6 Evolution of Preferences

In this section we describe how to enrich the basic model to deal with the evolution of preferences. The results of this section are not used later in the paper (and readers not interested in the evolution of preferences may skip to the next section.)

## 6.1 The Existing Literature

Starting with the seminal paper of Guth & Yaari (1992) (see also early proponents of the idea in Becker 1976; Frank 1987), there have been several papers studying the endogenous evolution of preferences using the so-called "indirect evolutionary approach" (see, e.g., Ok & Vega-Redondo 2001; Sethi & Somanathan 2001; Dekel *et al.* 2007; Heifetz *et al.* 2007; Herold & Kuzmics 2009.) These papers consider populations in which agents may have subjective preferences that differ from the material payoffs, and they observe with positive probability the subjective preferences of the opponent,[14] and players learn to play a Bayesian-Nash equilibrium with respect to their subjective preferences. In a stable state, all incumbent preferences achieve the same fitness payoff, and any sufficiently small group of mutants with different subjective preferences is outperformed.

The standard argument for observing preferences is that people give signals that provide clues as to their feeling (e.g.., a blush may reveal a lie). As discussed in Robson & Samuelson (2010, Section 2.5), the emission of such signals and their correlation with preferences are themselves the product of evolution, and thus it is not clear what prevents the appearance of a mimic who emit the signal without having the associated preferences. Robson & Samuelson (2010) summarize their discussion by suggesting that "the indirect evolutionary approach will remain incomplete until the evolution of preferences, the evolution of signals about preferences, and the evolution of reactions to these signals, are all analyzed within the model."

## 6.2 Incorporating Endogenous Preferences in the Model

In this subsection we present a simple extension to the basic model[15] that includes subjective preferences and analyze the evolution of preferences, signals about preferences and reactions to these signals. The modeling of preferences and stability of preferences is done in an analogous way to Dekel *et al.* (2007), with one key difference: players observe past behavior rather then observing preferences directly.

[14] See Heller & Mohlin (2014) for a model in which observation and deception about the opponent's preferences and endogenously determined by the players' cognitive levels.

[15] One can also extend the general model in a similar way.

Let $\Theta = [0,1]^{|A|^2}$ be the set of all possible utility functions on $A \times A$ (modulus affine transformations). A typical element $\theta \in \Theta$ describes a *preference type* (or *type*). A *population state* (or *state*) is a pair $\left(q, S_q = (s_\theta)_{\theta \in C(q)}\right)$ where: (1) $q \in \Delta(\Theta)$ is a distribution (with a finite support) over the set of types, (2) $s_\theta \in S$ is the policy of type $\theta$. To simplify the notation, we assume that different types use different policies (the results are similar without this assumption). Given a state, let $\sigma(q, S_q)$ be the strategy that is induced by $q$ and $S_q$ (i.e., $(\sigma(q, S_q))(s_\theta) = q(\theta)$ for each $\theta \in C(q)$), and let $\eta(q, S_q)$ be the unique outcome consistent with the strategy $\sigma(q, S_q)$.

The observation structure $(p, \delta, \tilde{\alpha})$ remains as in the basic model. A population state is *preferences-consistent* (given the observation structure) if each type maximizes his subjective preferences; i.e., for each $\theta \in C(q)$ and each message $m \in M$ :

$$s_\theta(m) \in \text{argmax}_{a \in A} \left( \sum_{a' \in A} (\theta(a, a') \cdot \Pr(a'|m)) \right),$$

where $\Pr(a'|m)$ is the probability that the opponent plays $a'$ conditional on observing signal $m$ (given the population state). An explicit expression for $s_\theta(m)$ is the following (Where $\sigma = \sigma(q, S_q)$ is the strategy induced by the state and $\eta$ its unique outcome):

$$m \neq \phi \Rightarrow s_\theta(m) \in \text{argmax}_{a \in A} \left( \sum_{\theta' \in C(q)} \sigma(\theta') \cdot \eta_{s_{\theta'}, \sigma}(m) \cdot \left( \sum_{a' \in A} \eta_{s(\theta')}(s_\theta)(a') \cdot \theta(a, a') \right) \right),$$

$$\text{and } s_\theta(\phi) \in \text{argmax}_{a \in A} \left( \sum_{\theta' \in C(q)} \sigma(\theta') \cdot \left( \sum_{a' \in A} \eta_{s_{\theta'}}(s_\theta)(a') \cdot \theta(a, a') \right) \right).$$

That is, players play a Bayesian Nash equilibrium given their subjective preferences: Each player best-replies (maximizes his subjective preferences) to the Bayesian inference about his opponent's type that is obtained from the observed message.

Given two distributions of types $q^*, q'$ and $\epsilon > 0$, let $q_{q^*,\epsilon,q'}$ be the mixture distribution: $q_{q^*,\epsilon,q'} = (1-\epsilon) \cdot q^* + \epsilon \cdot q'$. Given two states $(q, S_\theta)$ and $(q', S'_\theta)$ such that $C(q) \subseteq C(q')$, we say that state $(q', S'_\theta)$ is *focal* with respect to $(q, S_\theta)$ if the two profiles induce the same strategies over $C(q)$, that is: $\theta \in C(q) \Rightarrow s_\theta = s'_\theta$.[16]

A state is neutrally stable if any sufficiently small group of mutants is weakly outperformed in any focal post-entry state (and such focal post-entry states exist.) Formally:

[16] Our notion of focality is somewhat stronger than Dekel *et al.* (2007)'s focality by requiring incumbents to play the same also after signals that have not been observed in the pre-entry state (the difference exists only in noiseless environments).

**Definition 4.** A consistent state $(q^*, S^*_\theta)$ is neutrally stable if there exists $\bar{\epsilon} > 0$ such that for each $0 < \epsilon < \bar{\epsilon}$ and each "mutant" distribution of types $q'$:

1. There exists a consistent focal "post-entry" state $\left(q_{q^*,\epsilon,q'}, \tilde{S}_q\right)$.
2. For any consistent and focal state $\left(q_{q^*,\epsilon,q'}, \tilde{S}_q\right)$ with an induced strategy $\tilde{\sigma} = \sigma\left(q_{q^*,\epsilon,q'}, \tilde{S}_q\right)$, the mutants are (weakly) outperformed: $\pi_{\sigma\left(q',\tilde{S}_q\right)}\left(\tilde{\sigma}\right) \leq \pi_{\sigma\left(q,\tilde{S}_q\right)}\left(\tilde{\sigma}\right)$.

The focus on focal post-entry states is motivated (similar to Dekel *et al.*, 2007) by the view that in a stable state the behavior of the incumbents should not change due to the presence of a small group of mutants.[17] We mainly consider neutral (rather than evolutionary) stability in this setup because "equivalent" types that have slightly different preferences but same observable behavior, will always get the same payoff as the incumbents.

## 6.3 Discussing Our Modeling Approach

The agents in our model behave as if they follow a "revealed preferences" approach: they assess the opponent's preference type by observing his past behavior. Unlike tightening of the lips, past behavior is a reliable cue about the opponent's preferences because it is costly to mimic. The only way in which Type $\theta'$ can mimic the signals of type $\theta$, is by presenting a similar behavior.

Our approach (in which preferences are observed by their influence on past behavior) also solves additional problems of the existing literature (in which preferences are observed directly):

1. If type $\theta'$ has slightly different preferences than $\theta$ without any real influence on the game (e.g., slight different numbers in the subjective payoff matrix, but the same best-reply correspondence given the population state), then in the existing literature it induces completely different signals. We think that it is implausible to assume that tightening of the lips (say) reveal slight numerical differences in the payoff matrix with no direct behavioral impact.

2. Various results in the existing literature crucially depend on non-generic subjective preferences, such as: (1) subjective preferences in which the agent is completely indifferent between all the action profiles (e.g., Dekel *et al.* 2007, Prop. 2), and (2) subjective preferences in which one action is subjectively weakly dominated by another action (e.g.,Dekel *et al.* 2007, Prop. 3.) This dependency on non-generic preferences do not occur in our setup.

[17] Similar, to Dekel *et al.* (2007), we could slightly weaken Definition 4, and allow a state to be stable also in there are no consistent focal states, but for each $\delta > 0$ there exist consistent states in which the incumbents' strategy change by at most $\delta$ (and in all such states the mutants are outperformed.) All the results discussed below remain the same with this weakening.

## 6.4 Results with endogenized Preferences

In what follows we show how the results of our model can be extended to the setup with endogenized preferences. The first result adapts Prop. 2 and shows that if the underlying game admits a strict equilibrium, then material preferences are stable. This is formalized as follows. Let $\pi$ denote the subjective preference type that coincides with the objective payoffs, and the degenerate distribution of types that assigns mass one to the objective payoffs.

**Proposition 4.** *Assume that $a^*$ is a strict equilibrium of the underlying game. Then the state $(\pi, a^*)$ is neutrally stable.*

The proof is omitted as it is almost identical to the proof of Prop. 2. Prop. 4 implies that the material preferences are stable in all games that admit strict equilibria. The result can be extended the general model in which players observe several actions and action profiles.

Prop. 4 contradicts a main stylized result in the existing literature according to which if the observation probability is sufficiently close to one, then only efficient outcomes (which maximize the sum of fitness payoffs) can be stable (see, e.g., Dekel *et al.* 2007, Prop. 7.) If one accepts our approach that preferences are observed by their influence on behavior, then observability of preferences is not enough to destabilize inefficient strict Nash equilibria. Such destabilization requires to enrich the model with additional components, such as: (1) assortative (non-uniform) matching (see, e.g., Alger & Weibull, 2013), (2) cheap-talk with unused signals that can be used as "secret-handshakes" among mutants (see, e.g., Robson, 1990), and (3) non-payoff monotonic selection dynamics (see, e.g., Frenkel *et al.*, 2014.)

*Remark* 2. One can show that if there is small positive level of noise ($\delta > 0$), then the set of states in which all the incumbents: (1) have preferences for which $a^*$ is a best-reply to itself, and (2) play the policy $a^*$, is an evolutionarily stable set a la Thomas (1985).

The second result adapts Prop. 3 and shows that in noisy environments defection is the unique stable outcome in the Prisoner's Dilemma, and that only preferences for which defection is the best-reply to itself can be stable.

**Proposition 5.** *Assume that $G$ is a separable Prisoner's Dilemma game, that $\delta > 0$, and that $\left(q^*, S_q^*\right)$ is a neutrally stable state with strategy $\sigma^*$. Then $\sigma^* \equiv d$, and all preferences in $C\left(q^*\right)$ satisfy that $d$ is a best-reply to itself.*

One can also extend the general model to a setup with endogenous preferences, and show that when players observe action profiles (rather than actions), then cooperation can be a stable outcome and that it is supported by agents with non-material coordinating preferences ($c$ is a strict best reply to $c$, and $d$ is a strict best reply to $d$.)

# 7 General Model (Observing Multiple Actions)

We extend the basic model to deal with observation of several past actions, and action profiles.

## 7.1 Signals and Strategies

Before playing the game, each player with probability $0 \leq p_1 \leq 1$ privately observes $k \geq 1$ actions played in the past by his opponent, and with probability $0 \leq p_2 \leq 1 - p_1$ privately observes $k \geq 1$ action profiles played in past interactions of his opponent. With the remaining probability $p_0 \equiv 1 - p_1 - p_2$, the player observes a non-informative signal $\phi$.

We will refer to the tuple of parameters $(p_1, p_2, k)$ as the *observation structure* of the environment. Let the set of messages be adapted to the general model as follows: $M = \phi \cup A^k \cup (A \times A)^k$.

*Remark* 3. The assumption that the number of observation is fixed (either 0 or $k$) is made only to simplify the notations. All of the results can be extended to a setup in which the number of observation has a larger support.

## 7.2 Configurations

The definitions of a policy, a strategy and an outcome remains the same as in the basic model. Recall that $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ is the mapping between outcomes that is induced by $\sigma$ (that is, $f_\sigma(\eta)$ is the new outcome induced by players who follow strategy $\sigma$, and before playing, may observe independent realizations of the opponent's past behavior according to the outcome $\eta$.) We adapt the formal definition of $f_\sigma$ to the general setup as follows:

$$\begin{aligned}\left(f_\sigma(\eta)\right)_s(s')(a) &= p_0 \cdot s(\phi)(a) + p_1 \cdot \sum_{(a_i)_{i\leq k_1} \in A^{k_1}} \prod_{i \leq k} \eta_{s',\sigma}(a_i) \cdot s\left((a_i)_{i\leq k_1}\right)(a) \\ &\quad + p_2 \cdot \sum_{(a_i,a_i')_{i\leq k_2} \in A\times A} \prod_{i\leq k} \mu_{s',\sigma,\eta}(a_i, a_i') \cdot s\left((a_i, a_i')_{i\leq k_2}\right)(a).\end{aligned}$$

In this general setup, it is still true that any strategy admits a consistent outcome (Lemma 1), but, unlike the basic model, this consistent outcome does not have to be unique (even when $p < 1$) as demonstrated in the following example.

**Example.** Assume that the underlying game has two actions $A = \{c, d\}$, and that the information structure is: $p_1 = 0.8$, $p_2 = 0$ and $k = 3$. Let $\tilde{s}$ be the following "majority-tit-for-tat" policy: each action is chosen with equal probability if no past actions are observed ($\tilde{s}(\phi)(c) = 0.5$), and

otherwise the player plays the action his partner played in at least two of the three observed actions. Note that outcome $\eta \in O_{\{\tilde{s}\}} \equiv \Delta(A)$ is consistent with the strategy $\tilde{s}$ iff $\eta(c)$ solves the following equation: $x = 0.2 \cdot 0.5 + 0.8 \cdot (x^3 + 3 \cdot x^2 \cdot (1-x))$. Simple algebraic calculations show that the equation has three solutions, which are (approximately): 0.15, 0.5 and 0.85.

Thus, we have to describe the behavior of a population by a configuration - a pair consisting of a strategy and a consistent outcome. Formally:

**Definition 5.** A *configuration* is a pair $(\sigma, \eta)$, where $\sigma \in \Sigma$ is a strategy and $\eta \in O_{C(\sigma)}$ is a consistent outcome (i.e., $f_\sigma(\eta) \equiv \eta$.)

The payoff of a policy depends now also on the outcome. Thus, given a configuration $(\sigma, \eta)$ and a policy $s \in C(\sigma)$, we adapt the definition of the payoff of a policy as follows:

$$\pi_s(\sigma, \eta) = \sum_{(a,a') \in A \times A} \pi(a, a') \cdot \mu_{s,\sigma,\eta}(a, a'),$$

and given a strategy $\sigma'$ with a smaller support than $\sigma$ ($C(\sigma') \subseteq C(\sigma)$), let $\pi_{\sigma'}(\sigma, \eta)$ be:

$$\pi_{\sigma'}(\sigma, \eta) = \sum_{s' \in C(\sigma')} \sigma'(s') \cdot \pi_{s'}(\sigma, \eta).$$

We say that a configuration is *balanced* if $\pi_s(\sigma, \eta) = \pi_{s'}(\sigma, \eta)$ for every $s, s' \in C(\sigma)$, $\pi_s(\sigma, \eta) = \pi_{s'}(\sigma, \eta)$; in that case, we write the uniform payoff as $\pi(\sigma, \eta)$. The definition of a post-entry strategy is unchanged. A post-entry configuration is a pair consisting of a post-entry strategy and a consistent outcome.

# 8 Solution Concepts for the general Model

The fact that the outcome is not uniquely determined by the strategy, requires us to adapt the definition of evolutionary stability. We do this adaptation in Subsection 8.1 and discuss its implication in Subsection 8.2. In the remaining two subsections, we adapt to the current setup Selten (1983)'s notion of limit ESS, to deal with different kinds of rare noise and rare mistakes.

## 8.1 Evolutionary stability

A configuration is evolutionary (neutrally) stable if any sufficiently small group of mutants is strictly (weakly) outperformed, and in addition, their influence on the outcome is small. The stability is weak if we require at least one post-entry outcome to have these properties and it is

strong if we require all post-entry configurations to have these properties. The weak definition is used for the uniqueness results, while the strong definition is used for existence results. Formally:

**Definition 6.** Configuration $(\sigma^*, \eta)$ is *strongly evolutionary (neutrally) stable* if for any $\delta > 0$, there exists $\bar{\epsilon} > 0$ such that for each $0 < \epsilon < \bar{\epsilon}$, each "mutant" strategy $\sigma' \in \triangle(B)$, and each post-entry configuration $(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta})$:

1. The post-entry outcome is close: For each $s \in C(\sigma)$ and actions $a, a' \in A$, the following inequality holds: $|\mu_{s,\sigma,\tilde{\eta}}(a, a') - \mu_{s,\sigma,\tilde{\eta}}(a, a')| < \delta$; and
2. The incumbents outperform the mutants:

$$\sigma \neq \sigma' \Rightarrow \pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta}) < \pi_{\sigma}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta}) \quad (\pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta}) \leq \pi_{\sigma}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta})).$$

**Definition 7.** Configuration $(\sigma^*, \eta^*)$ is *weakly evolutionary (neutrally) stable* if for any $\delta > 0$, there exists $\bar{\epsilon} > 0$ such that for each $0 < \epsilon < \bar{\epsilon}$, and each "mutant" strategy $\sigma' \in \triangle(B)$, there exists a post-entry configuration $(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta})$ with a nearby outcome:

1. For each $s \in C(\sigma)$ and $a, a' \in A$ the following holds: $|\tilde{\eta}_s(a, a') - \eta^*_s(a, a')| < \delta$; and
2. the incumbents in this configuration outperform the mutants:

$$\sigma \neq \sigma' \Rightarrow \pi_{\sigma'}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta}) < \pi_{\sigma}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta}) \quad (\pi_{\sigma'}(\sigma\sigma^*{}_{,\epsilon,\sigma'}, \tilde{\eta}) \leq \pi_{\sigma}(\sigma_{\sigma^*,\epsilon,\sigma'}, \tilde{\eta})).$$

## 8.2 Observations on Evolutionarily Stable Configurations

In this subsection we present simple observations about the notions of stability defined above.

First, observe that there are immediate logic relations between the different notions: strong evolutionary stability implies all the notions, and weak neutral stability is implied by all the notions (see also Figure 1 at the end of the section.) Second, note that our definitions coincide with the standard notions of evolutionary stability (Maynard-Smith, 1974) when there is no observability (because without observability a policy is mixed action, and the consistent outcome is the profile of mixed actions).

**Fact 1.** *Assume that $p_0 = 1$ (i.e, no observability of the opponent's past actions.) Then:*

1. *Configuration $(\sigma^*, \eta^*)$ is strongly evolutionary stable $\Leftrightarrow$ $(\sigma^*, \eta^*)$ is weakly evolutionary stable $\Leftrightarrow$ strategy $\sigma^*$ is evolutionary stable in the underlying game.*

2. *Configuration* $(\sigma^*, \eta^*)$ *is strongly neutrally stable* $\Leftrightarrow$ $(\sigma^*, \eta^*)$ *is weakly neutrally stable* $\Leftrightarrow$ *strategy* $\sigma^*$ *is neutrally stable in the underlying game.*

Next, note that any weakly neutrally stable configuration is balanced.

*Claim* 3. Let $(\sigma^*, \eta^*)$ be weakly neutrally stable configuration. Then $(\sigma^*, \eta^*)$ is balanced.

*Proof.* Assume to the contrary that there is a policy $s'$ such that $\pi_{s'}(\sigma^*, \eta^*) > \pi_{\sigma^*}(\sigma^*, \eta^*)$. Let $0 < \delta << \pi_{s'}(\sigma^*, \eta^*) - \pi_{\sigma^*}(\sigma^*, \eta^*)$. The fact that $(\sigma^*, \eta^*)$ is weakly neutrally stable implies that there exists $\bar{\epsilon} > 0$ such that for each $0 < \epsilon < \bar{\epsilon}$ there exists a configuration $(\sigma_{\sigma^*,\epsilon,s'}, \tilde{\eta})$such that $|\tilde{\eta}_{s'}(a, a') - \eta^*_{s'}(a, a')| < \delta$. The previous inequalities implies that $\pi_{s'}(\sigma_{\sigma^*,\epsilon,\sigma'}, \eta) > \pi_{\sigma^*}(\sigma_{\sigma^*,\epsilon,\sigma'}, \eta)$, which contradicts the second inequality implied by weak neutral stability (that the incumbents should weakly outperform the mutants.) □

The additional requirement that a small invasion of mutants will have a small influence on the outcome is an adaptation of Dekel *et al.* (2007)'s notion of stability, and its focus on focal post-entry configurations. Our final observation shows that strong stability implies unique outcome.

*Claim* 4. Let $(\sigma^*, \eta^*)$ be strongly neutrally stable configuration. Then if $(\sigma^*, \eta')$ is configuration it implies that $\eta^* = \eta'$.

*Proof.* Assume to the contrary that $(\sigma^*, \eta^*)$ is a strongly neutrally stable configuration, and that $(\sigma^*, \eta')$ is a configuration and $\eta^* \neq \eta'$. Let $0 < \delta < \max_{s \in C(\sigma), a, a' \in A} |\eta'_s(a, a') - \eta^*_s(a, a')|$. Consider a mutant strategy $\sigma' = \sigma^*$. It is immediate that for each $\epsilon > 0$ the post-entry configuration $(\sigma_{\sigma^*,\epsilon,\sigma'}, \eta') = (\sigma^*, \eta')$ is not nearby (violates condition (1) of the definition of strong neutral stability), and we get a contradiction. □

## 8.3 Limit Evolutionary Stability

Our strategic interaction has two stages: observing signals about the opponent, and playing an action in the underlying game. As common in interactions with multiple stages, evolutionary stability is too demanding in our setup: typically not all signals about the opponent are observed on the equilibrium path, and as a result there exist strategies that are realization equivalent to the incumbent strategy and differ only in actions following the observation of a signal off the equilibrium path. This motivates us to slightly weaken the notion of evolutionary stability (a la Selten, 1983's notion of limit ESS as reformulated in Heller, 2014) and require evolutionary stability only in a converging sequence of perturbed games in which players rarely "tremble" and choose a "wrong" policy or a "wrong" action (but not necessarily in the unperturbed game). As discussed

later (Remark 5), trembling when choosing actions is closely related to noisy observations of the basic model.

We begin by defining a perturbation in our environment.

**Definition 8.** A *perturbation is a tuple* $\zeta = (\xi, \mathcal{S}, \lambda)$ where:

1. $\xi : A \to \mathbb{R}^+$ is a function that assigns a non-negative number for each action such that $\sum_{a \in A} \xi(a) < 1$ .
2. Let $S_\xi \subseteq S$ the set of policies that assign probability of at least $\xi(a)$ to each action $a$ after any observed signal. $\mathcal{S} \subseteq S_\xi$ is a finite set of policies.
3. $\lambda : \mathcal{S} \to \mathbb{R}^+$ is a function that assigns a non-negative number for each policy in $\mathcal{S}$ such that $\sum_{s \in \mathcal{S}} \lambda(s) < 1$.

Let $\mathbf{\Gamma}(\zeta) = \mathbf{\Gamma}(\xi, \mathcal{S}, \lambda)$ denote the *perturbed environment* that results from perturbing the environment described in Section 2 by perturbation $\zeta$. Let the convex set of feasible strategies $\Sigma(\zeta) \subseteq \Sigma$ in environment $\mathbf{\Gamma}(\zeta)$ be defined as follows. Strategy $\sigma \in \Sigma$ is included in $\Sigma(\zeta)$ iff: (1)$C(\sigma) \subseteq S_\xi$, and (2) For each $s \in \mathcal{S}$, $\sigma(s) \geq \lambda(s)$ . That is, in $\mathbf{\Gamma}(\zeta)$ players tremble both when being assigned into a policy and when choosing an action given the policy: each strategy must assign probability of at least $\lambda(s)$ to policies in $\mathcal{S}$, and each policy in the support of the strategy must assign probability of at least $\xi(a)$ to each action $a$. Let $L(\zeta)$ denote the *maximal tremble* of $\zeta = (\xi, \mathcal{S}, \lambda)$: $L(\zeta) = \max(\max_{a \in A} \xi(a), \max_{s \in \mathcal{S}} \lambda(s))$ .

*Remark* 4. Note that trembles when choosing the policy are closely related to the trembles used in the notion of normal-form perfection (Selten (1975)), while trembles when choosing actions are equivalent to the more "common" trembles used in the definitions of extensive-form perfection and limit ESS (Selten, 1983). See Selten (1975, Section 13) and van Damme (1987, Section 6.4) for discussing the differences between normal-form perfection in which players tremble when choosing a strategy for the entire game (a policy in our setup), and extensive-form perfection in which players tremble when choosing actions at each information set.

A limit evolutionary stable configuration is the limit of evolutionary stable configurations in a converging sequence of perturbed games. Formally:

**Definition 9.** Configuration $(\sigma^*, \eta^*)$ is *limit strongly (weakly) evolutionary stable* if for any $\delta > 0$, there exists $\bar{\epsilon} > 0$ (called, invasion barrier) and a sequence $(\zeta_n, (\sigma_n, \eta_n))$ such that:

1. $\zeta_n$ is a sequence of perturbations converging to the unperturbed game: $L(\zeta) \to 0$; and

2. Each strategy $\sigma_n$ is feasible in the perturbed game $\zeta_n$: $\sigma_n \in \Sigma(\zeta_n)$, and each pair $(\sigma_n, \eta_n)$ is a configuration. The sequence of configurations $(\sigma_n, \eta_n)$ converges to $(\sigma^*, \eta^*)$; and

3. For each $0 < \epsilon \leq \bar{\epsilon}$, each $n \in \mathbb{N}$, each feasible mutant strategy $\sigma'_n \in \Sigma(\zeta_n)$, and each (there exists a) post-entry configuration $\left(\sigma_{\sigma_n,\epsilon,\sigma'_n}, \tilde{\eta}\right)$ (such that) the following hold :

   (a) Focality: For each $s \in C(\sigma_n)$ and actions $a, a' \in A$: $|(\tilde{\eta}_n)_s(a, a') - (\eta_n)_s(a, a')| < \delta$.

   (b) The incumbents outperform the mutants: $\sigma_n \neq \sigma'_n \Rightarrow \pi_{\sigma'_n}(\sigma_{\sigma_n,\epsilon,\sigma'}, \tilde{\eta}) < \pi_{\sigma_n}(\sigma_{\sigma_n,\epsilon,\sigma'}, \tilde{\eta})$.

The stability of a limit evolutionary stable configuration depends on the structure of the "trembles" - the relative frequency of different "mistakes" that the incumbents do. That is, such a strategy is a limit of evolutionarily stable configurations for a given converging sequence of perturbed games, but it may no be such a limit with respect to a different converging sequence of perturbed games in which other trembles are more frequent. Motivated by this observation and by the assumption that in many setups trembles are more likely to appear when choosing actions (rather than when choosing policies), we present a stronger definition that requires stability with respect to a large set of perturbations: all perturbations in which players tremble when choosing actions (but not when choosing policies), and these trembles have full support. Formally:

**Definition 10.** A *perturbation* $\zeta = (\xi, \mathcal{S}, \lambda)$ is an *action perturbation with full support* if:

1. Trembles with full support when choosing actions: $\xi(a) > 0$ for each action $a \in A$.

2. No trembles when choosing policies: $\lambda \equiv 0$.

We identify an action perturbation with full support with its first component $\xi$.

**Definition 11.** Configuration $(\sigma^*, \eta^*)$ is *strict limit strongly evolutionary stable* if for any $\delta > 0$, there exists $\bar{\epsilon} > 0$, such that for any converging sequence of action perturbations with full support $(\xi_n)_{n\in\mathbb{N}}$ (i.e., $L(\xi) \to 0$ ), there exists $n_0 \in \mathbb{N}$ and a sequence $(\sigma_n, \eta_n)_{n\geq n_0}$ satisfying:

1. Each strategy $\sigma_n$ is feasible in the perturbed game $\zeta_n$: $\sigma_n \in \Sigma(\zeta_n)$. The sequence of configurations $(\sigma_n, \eta_n)_{n\geq n_0}$ converge to $(\sigma^*, \eta^*)$; and

2. For each $0 < \epsilon < \bar{\epsilon}$, each $n \geq n_0$, each feasible mutant strategy $\sigma'_n \in \Sigma(\zeta_n)$, and each post-entry configuration $\left(\sigma_{\sigma_n,\epsilon,\sigma'_n}, \tilde{\eta}\right)$the following hold :

   (a) Focality: For each $s \in C(\sigma_n)$ and actions $a, a' \in A$: $|(\tilde{\eta}_n)_s(a, a') - (\eta_n)_s(a, a')| < \delta$.

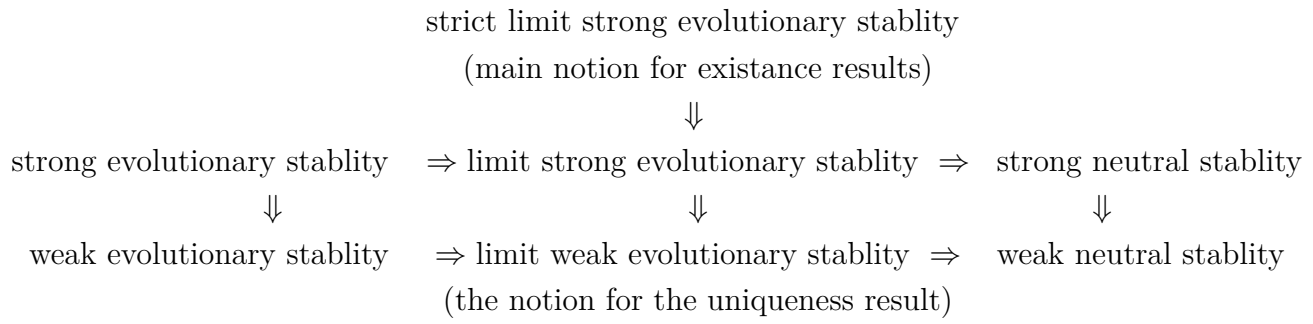
   (b) The incumbents outperform the mutants: $\sigma_n \neq \sigma'_n \Rightarrow \pi_{\sigma'_n}\left(\sigma_{\sigma_n,\epsilon,\sigma'_n}, \tilde{\eta}\right) < \pi_{\sigma_n}\left(\sigma_{\sigma_n,\epsilon,\sigma'_n}, \tilde{\eta}\right)$.

*Remark* 5. Action perturbations with full support have essentially the same influence on stability as the noise in the basic model: in both cases a player may observe any action with positive probability, and observing a "trembled"/"noisy" action does not influence the expected behavior of the opponent in the current interaction. There is a small difference between the two modeling approached: that when a player trembles, it slightly influences his payoff (while a noisy action is only observed by mistake and never actually played.) However, it turns out that this small difference do not affect the stability analysis in our setup. Specifically, one can show that strict limit evolutionary stability in the general model coincides with evolutionary stability with respect to any sufficiently small positive level of noise in the basic model.

## 8.4 Observations on Limit Evolutionary Stability

Figure 1 presents the logic relations between the different notions (which are immediate.)

Fig. 1: The Logic Relations Between the Different Solution Concepts.




The following simple observation (immediate from the definition) states that in a setup without observability, our definitions coincide with existing notions of stability:

uniform limit ESS (Heller, 2014), which slightly refines Selten's (1983) limit ESS,[18] and strict limit ESS (Heller (forthcoming), see also the related notion of strict perfection in Okada, 1981.)

**Fact 2.** *Assume that* $p_0 = 1$. *Then:*

[18] Heller (2014) shows that the original notion of limit ESS (Selten, 1983) is too weak: (1) it does not imply neutral stability, and (2) it may be dynamically unstable in the sense that almost any small perturbation takes the population away. These two issues are caused by the implicit assumption of the notion of "limit ESS" that mutants are more rare than "trembling" incumbents. Heller (2014) solves these two issues by defining a slightly stronger notion, *uniform limit ESS*, which requires mutants to be strictly outperformed also without this implicit assumption. All of our results remain the same if one adapts our definition of limit evolutionary stability by removing the requirement of uniform barrier as in the original definition of Selten (1983).

1. *Configuration* $(\sigma^*, \eta^*)$ *is strict limit strongly evolutionary stable* $\Leftrightarrow$ *strategy* $\sigma^*$ *is strict uniform limit evolutionary stable (Heller (forthcoming)) in the underlying game.*

2. $(\sigma^*, \eta^*)$ *is limit strongly evolutionary stable* $\Leftrightarrow$ $(\sigma^*, \eta^*)$ *is limit weakly evolutionary stable* $\Leftrightarrow$ $\sigma^*$ *is uniform limit evolutionary stable (Heller (2014)) in the underlying game.*

# 9 Results in the General Model

In this section we extend and strengthen the various results obtained previously for the special case of observing a single action.

## 9.1 Strict Stability of Strict Equilibria

In this subsection we show that each strict equilibrium $a^*$ is the pure outcome of a strict limit evolutionary stable configuration (which generalizes and strengthens Prop. 2.) The intuition is as follows. Consider a slightly perturbed game in which players rarely tremble when choosing actions. In such perturbed games, all signals are observed with positive probability. Consider a configuration in which all the (non-trembling) incumbents play the equilibrium action $a^*$ regardless of the signal about the opponent. Observe that any mutant strategy plays $a^*$ less often against the incumbents, and thus yields a strictly lower payoff in such interactions. If the mutants are sufficiently rare, then this loss cannot be compensated by a gain when facing other mutants. This implies that the configuration in which everyone always plays $a^*$ is strong limit evolutionary stable. Formally (proof is in the appendix):

**Definition 12.** Let $(\sigma, \eta)$ be a configuration. If there exists an action $a^*$ such that $\eta \equiv a^*$ (i.e., $\eta_s(s')(a^*) = 1$ for each $s, s' \in C(\sigma)$), then we say that the configuration is *pure* and write it as $(\sigma, a^*)$. Moreover, if the strategy includes a single policy that chooses action $a^*$ regardless of the signal about the opponent, we write it as $(a^*, a^*)$.

**Theorem 1.** *If action* $a^*$ *is a strict equilibrium then the pure configuration* $(a^*, a^*)$ *is a strict limit strong evolutionary stable configuration.*

*Remark* 6. Theorem 1 shows that strict equilibria are stable if players tremble when choosing actions. The stability may not hold in setups in which players mainly tremble when choosing policies, as demonstrated in an underlying $2 \times 2$ coordination game in which the players obtain payoff of two (one) if they both play $a$ (b), and they get payoff of zero if they choose different actions. One can show that the efficient action $a$ is stable with respect to all trembles. However, the stability of the inefficient action $b$ may fail if the more likely trembles are those in which

trembling players choose a policy that keep playing the same pure action. In such perturbed games there is no stable configuration in which players play $b$ with high probability. This is because if a player observes his opponent to play only $a$ in the past, then the unique best reply is to play $a$, and this allows a mutant who always play $a$ to invade the population.

## 9.2 Revisiting the Prisoner's Dilemma

In this subsection revisit the Prisoner's Dilemma with the more general observation structure. Our first result generalizes Section 3's result by showing the mutual defection is the unique stable outcome also if players observe several actions. Our second result show that mutual cooperation can be stable if players observe action profiles.

### 9.2.1 Defection is the Unique Stable Outcome ($p_2 = 0$)

In this subsection we restrict attention to a setup in which players play separable Prisoner's Dilemma and they only observe the opponent's past actions (rather than action profiles.) Under these assumptions we obtain a sharp uniqueness result: defection is the unique stable outcome. Formally:

**Theorem 2.** *Assume that $G$ is a separable Prisoner's Dilemma game, and that $p_2 = 0$. If $(\sigma, \eta)$ is a limit weak evolutionary stable configuration, then $\eta \equiv d$.*

The intuition of the result is similar to the one given earlier for Proposition 3 in Section 4.2. The formal proof appears in Appendix A.4.

*Remark* 7. The difficulty in supporting cooperation as a stable outcome in the Prisoner's Dilemma is induced by the fact that defection is both (1) a profitable deviation from cooperation, and (2) the unique way to punish deviations. If one enriches the Prisoner's Dilemma by adding a third action $e$ that satisfies: (1) $e$ is a best-reply to $d$, (2) $e$ "punishes" defectors $\pi(d, e) < \pi(c, c)$ , and (3) $e$ is not a profitable deviation form cooperation - $\pi(e, c) < \pi(c, c)$, then one can support cooperation as a stable outcome as follows: the incumbents cooperate when observing cooperation (or $\phi$), and play $e$ when observing at least one past defection (which is supported by a policy-perturbation in which trembling players choose a policy according to which they defect with probability close to one.) In this setup, one can also support a stable heterogeneous population in which the population includes two fractions: (1) one fraction always cooperate, and (2) the other fraction (Defectors) plays $d$ when observing a cooperator, and play $e$ when observing a defector (see Heller (forthcoming) for a related heterogeneous stable population in a setup in which agents with limited foresight play the repeated Prisoner's dilemma.)

*Remark* 8. Theorem 2 is related to an existing result of Takahashi (2010, Prop. 1), which deals with a setup in which a player can observe the entire history of past actions of his current partner. Takahashi shows that if $g \geq l$ then there is no strict equilibrium (i.e., an equilibrium in which each player after each message strictly prefers the action induced by the equilibrium) other then defection. Our result is stronger (though the setups are somewhat different) in that we show the uniqueness of defection for a weaker solution concept: neutrally stable strategy rather than the stronger notion of strict equilibrium. Takahashi (2010, Prop. 2) also presents an equilibrium that allow cooperation when $l > g$, in which each player is always indifferent between cooperation and defection; however, this equilibrium is not neutrally stable: the equilibrium is not robust to the presence of an arbitrary small group of "mutant" defectors.

### 9.2.2 Stable Cooperation when Observing action Profiles

In this subsection, we show that cooperation can be stable when players observe action profiles (rather than actions.) Specifically, Theorem 3 shows that observation of a single action profile can be enough to sustain cooperation as a limit strong evolutionary stable outcome.

**Theorem 3.** *Assume that the underlying game is Prisoner's Dilemma, the observation structure is* $(0, p_2, 1)$ *and* $g < p_2 < 1$. *Then there exists a strategy* $\sigma^*$ *such that the configuration* $(\sigma^*, c)$ *is a limit strong evolutionary stable.*

The sketch of proof is as follows (proof is in Appx. A.5.) Strategy $\sigma^*$ includes a single policy:

$$s^*(m) = \begin{cases} d & m = (d, c) \\ c & \text{otherwise} \end{cases}.$$

That is, the policy induces players to cooperate in all cases except when they observe that the opponent was the sole defector in the past; in this case they defect. We consider a converging sequence of perturbed games in which players rarely tremble by choosing with small probability the policy that always defects. Note that all action profiles are observed with positive probability in these perturbed games. Moreover, when an incumbent is observed to play $(d, c)$, it implies that he is a "trembler" who follows policy $d$ and thus is going to defect in the current interaction. Thus, mutants who cooperate instead of defecting after observing $(d, c)$ are strictly outperformed when facing the incumbents: they suffer an immediate loss of $g$, without gaining any indirect advantage from being observed to cooperate. Next, we show that mutants who always defect are strictly outperformed. Note that the probability that an incumbent defects when facing a mutant (which is the probability that a random past action profile of the mutant is equal to $(d, c)$) must be $p_2$

times the probability that an incumbent cooperates when facing a mutant. This implies that the incumbents defect with probability $\frac{p_2}{p_2+1}$ against the mutants. Thus the mutants payoff against the incumbents is given by: $\frac{1+g}{p_2+1}$, which is strictly smaller than the incumbents' payoff among themselves (which is equal to one) if $p_2 > g$. This implies that if the mutants are sufficiently rare, they are strictly outperformed.

*Remark* 9. The stability of cooperation relies on a particular kind of trembles, in which players are more likely to tremble when choosing policies. The stability does not hold in a setup in which players are more likely to tremble when choosing actions (and in particular, cooperation is not a *strict* limit evolutionary stable). In Section 10, we show that with public signals, cooperation becomes a *strict* limit evolutionary stable.

# 10 Public Signals

In the main model we assume that the signal about the opponent's behavior is private. In some applications it might be more plausible that the signals are public. In particular, if we consider an online interaction between traders through an intermediary web site that publicly presents feedback about the past behavior of the traders (e.g., the popular web site "eBay"), then in such interactions the signals about the past behavior (i.e., the trader's summary of feedback in the intermediary web site) are public. Another related setup in which public signals are more likely is an environment in which a player observes the last $k$ actions of the opponent, rather than $k$ random actions form the past. In such environments, the signals are essentially public because each player can remember his own recent history. In what follows we show how to adapt the model to public signals, and we analyze the influence that this change has on our results.

## 10.1 Changes to the Model

Before playing the game, each player *publicly* observes the signals about the opponent's past behavior. Because both players observe both signals, we redefine M as follows: $M = M_1 \times M_2 = \left(\phi \cup A^k \cup (A \times A)^k\right) \times \left(\phi \cup A^k \cup (A \times A)^k\right)$, where the first component is interpreted as the opponent's observed past behavior and the second component as the player's own observed past behavior. We denote a typical element of $M_1$ as $m_1$ (the message about the opponent's behavior), and a typical element of $M_2$ as $m_2$ (the message about the player's own behavior). All other details of the model remain the same. One can see that Theorems 1-2 remain the same with public signals, and the proofs require only minor adaptations.

## 10.2 Robust Stability of Cooperation with Public Signals

We now show that we can strengthen Theorem 3 in the setup of public signals, and show that cooperation is stable with respect to any sequence of action perturbations with full support (unlike the case of private signals, in which the stability relies on a specific sequence of policy perturbations.) Formally:

**Theorem 4.** *Assume that the underlying game is Prisoner's Dilemma, the observation structure is* $(0, p_2, 1)$ *with* $p_2 > g$*, and the signals are public. Then there exists a strategy* $\sigma^*$ *such that the configuration* $(\sigma^*, c)$ *is a strict limit strong evolutionary stable.*

The sketch of proof is as follows (proof is in Appx. A.6.) Strategy $\sigma^*$ includes a single policy:

$$s^*(m) = \begin{cases} d & m_1 = (d,c) \text{ or } m_2 = (d,c) \\ c & \text{otherwise} \end{cases}.$$

That is, the policy induces players to cooperate in all cases except when they observe that either of the players was the sole defector in the past; in this case they defect. We consider an arbitrary converging sequence of action perturbations with full support (i.e., players rarely tremble and play the "wrong" action with small positive probability.) Note that all action profiles are observed with positive probability in these perturbed games. Moreover, mutants who cooperate instead of defecting after observing $(d, c)$ are strictly outperformed when facing the incumbents: they suffer an immediate loss of $l$, without gaining any indirect advantage from being observed to cooperate (as the opponent is going to defect with probability very close to one). The argument why mutants who always defect are strictly outperformed is the same as the one given after Theorem 3. This implies that if the mutants are sufficiently rare, they are strictly outperformed.

*Remark* 10. The policy $s^*$ is closely related to the strategy "Pavlov" (AKA, "win-stay, lose-change") in the standard repeated Prisoner's Dilemma in which a player defects iff the players played different actions in the previous round (see, Kraines & Kraines, 1989; Nowak & Sigmund, 1993, andHeller, forthcoming.) "Pavlov" was used several times in the literature to support cooperation with direct reciprocity (players interact in repeated interactions, and each player rewards his past opponent's behavior), but to the best of our knowledge the current paper is the first to apply it to in a setup of indirect reciprocity (in which each pair of players meet only once).

# 11 Related Literature

In this section we discuss the relations between our paper and the related literature (excluding literature discussed elsewhere in the paper).

A few papers study how cooperation can be supported in the Prisoner Dilemma without observing the opponent's past behavior. Ellison (1994) (extending previous results of Kandori, 1992 and Harrington Jr, 1995) shows how cooperation can be supported in finite populations by "contagious" punishments (if one player defects at stage $t$, his partner defects from period $t+1$, infecting another player who defects from period $t+2$ on, etc.); however, such "contagious" punishments can only work in finite populations, and if the players are sufficiently patient with respect to the population size. Fujiwara-Greve & Okuno-Fujiwara (2009) show how cooperation can be supported in "voluntarily separable" repeated Prisoner's Dilemma, in which each player can unilaterally end and start with a randomly assigned new partner with no information flow.

Rosenthal (1979) presented an early model in which players in a population are randomly matched, and each player can observe his opponent's last action (See also Okuno-Fujiwara & Postlewaite, 1995.) A few papers constructed models in which players may observe the partner's planned action (rather than past action): Robson (1994) presented a model in which each player has small probability to observe his partner's planned action and revise his own action; Solan & Yariv (2004) dealt with a setup in which one of the players can spend effort and observe his opponent's action before playing.

## 12 Conclusion

In this paper we study a setup in which players are randomly matched, and each player may observe signals about the opponent's past behavior. We present three main results: (1) strict equilibria of the underlying game are stable outcomes, (2) defection is the unique stable outcome in the Prisoner's Dilemma when players only observe past actions, and (3) cooperation can be a stable outcome if players observe action profiles (and the stability becomes more robust if the signals are public.) Moreover, we show that our model can be extended to study endogenous observability (which depends on the players' efforts) and the evolution of subjective preferences.

Throughout the paper we interpreted the players in the model as naive agents who just follow their programmed strategies. However, similar to other applications of evolutionary stability a la Maynard Smith & Price (1973), the results are robust to the presence of sophisticated agents who explicitly maximize their payoff. Specifically, our results can be adapted to a setup a la Kandori (1992) in which sophisticated patient players in a large population are randomly matched at each round to play the underlying game.

In what follows we sketch three interesting directions for future research. In this paper we mainly applied our model to the Prisoner's Dilemma. It will be interesting to apply the model to other games, and in particular, to the Hawk-Dove game, in which we conjecture that one

can characterize a stable heterogeneous population in which "committed Hawks" and "flexible players" co-exist.

Second, our extension to subjective preferences is somewhat limited because players only interact in single game. It seems intriguing, to study richer environments in which players are endowed with "universal" (non-game-specific) preferences over fitness profiles, and they interact in different games (and use the same preferences in all these games). Another interesting direction (pursued in a companion working paper, Heller & Mohlin, 2014) is allowing agents to spend effort in deception - influencing the signal observed by the opponent.

Finally, our model assumes that players directly observe past actions of the partner. In many economic setups, it seems more plausible that agents only observe the reports of other agents about the past interactions of their partner (e.g., the trader's feedback profile in eBay.) A policy in this setup should specify both the played actions, as well as the reports to other agents, and it will be interesting to characterize stable outcomes in such setups.

# A Proofs

## A.1 Proof of Prop. 1 (Unique Consistent Outcomes)

*Proof.* Assume to the contrary that $\eta \neq \eta'$. Given two mixed actions $\alpha, \alpha' \in \Delta(A)$, define their distance as the sum of the differences in the weights they assign to the different actions:

$$\|\alpha - \alpha'\| = \sum_{a \in A} |\alpha(a) - \alpha'(a)|.$$

We interpret $0 \leq \|\alpha - \alpha'\| \leq 1$ as the probability in which $\alpha$ and $\alpha'$ induce different actions (in this interpretation we implicitly assume the maximal dependency between the two lotteries; thus if both lotteries have the same distribution, they always induce the same outcome.)

Given two outcomes $\eta, \eta' \in O_{C(\sigma)}$, define their distance as the maximal distance between the mixed actions $\eta_s(s'), \eta'_s(s')$ for each two policies $s, s' \in \sigma$:

$$\|\eta - \eta'\| = \max_{s,s' \in C(\sigma)} \|\eta_s(s') - \eta'_s(s')\|.$$

Observe that for any policy $s' \in C(\sigma)$ and any action $a \in A$:

$$\begin{aligned}
\left\|\eta_{s,\sigma} - \eta'_{s,\sigma}\right\| &= \left\| \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s') - \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta'_s(s') \right\| \\
&= \sum_{a \in A} \left| \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta_s(s')(a) - \sum_{s' \in C(\sigma)} \sigma(s') \cdot \eta'_s(s')(a) \right| \\
&= \sum_{a \in A} \left| \sum_{s' \in C(\sigma)} \sigma(s') \cdot (\eta_s(s')(a) - \eta'_s(s')(a)) \right| \\
&\leq \sum_{a \in A} \sum_{s' \in C(\sigma)} \sigma(s') \left| (\eta_s(s')(a) - \eta'_s(s')(a)) \right| \\
&= \sum_{s' \in C(\sigma)} \sigma(s') \sum_{a \in A} \left| (\eta_s(s')(a) - \eta'_s(s')(a)) \right| \\
&= \sum_{s' \in C(\sigma)} \sigma(s') \cdot \|\eta_s(s') - \eta'_s(s')\| \leq \max_{s' \in C(\sigma)} \|\eta_s(s') - \eta'_s(s')\| = \|\eta - \eta'\|;
\end{aligned} \tag{1}$$

where the equalities and the last inequality are immediately implied by the definitions above, while the first inequality is a triangle inequality ($|\sum_i x_i| \leq \sum_i |x_i|$).

Let $s, s' \in C(\sigma)$. Consider a player who follows policy $s$ , faces an opponent with policy $s'$ and observes a message according to $\eta$ (case I) or $\eta'$ (case II). Recall, that the expression $\|(f_\sigma(\eta))_s(s') - (f_\sigma(\eta'))_s(s')\|$ can be interpreted as the probability in which the observed action of the player differs in these two cases. The observed action differs between these two cases only if he (1) observes a past action (which occurs with probability $p < 1$), and (2) the observed message is different in the two cases (which happens with probability of $\left\|\eta_{s',\sigma} - \eta'_{s',\sigma}\right\|$). This implies that:

$$\|(f_\sigma(\eta))_s(s') - (f_\sigma(\eta'))_s(s')\| \leq p \cdot \left\|\eta_{s',\sigma} - \eta'_{s',\sigma}\right\| \leq p \cdot \|\eta - \eta'\| < \|\eta - \eta'\|; \tag{2}$$

where the second inequality is implied by (1) and the third inequality holds because $p < 1$.

Inequality (2) implies a contradiction:

$$\|\eta - \eta'\| = \|f_\sigma(\eta) - f_\sigma(\eta')\| = \max_{s' \in C(\sigma)} \|f_\sigma(\eta)(s') - f_\sigma(\eta')(s')\| < \|\eta - \eta'\|.$$

□

## A.2 Proof of Prop. 3 (Only Defection is Stable)

*Proof.* We first observe that in the separable Prisoner's Dilemma, the payoff of a policy depends only on its probability of defection. That is, for each strategy $\sigma$ (with unique consistent outcome

$\eta$) and policy $s \in C(\sigma)$ :

$$\begin{aligned}
\pi_s(\sigma) &= (\Pr(\text{opponent cooperates})) \cdot (1+g) - \eta_s(c) \cdot g \qquad (3)\\
&= \sum_{m \in M} (\Pr(\text{opponent observes } m\ ) \cdot \sigma_m(c)) \cdot (1+g) - \eta_s(c) \cdot g \\
&= (p \cdot (\eta_s(d) \cdot \sigma_d(c) + \eta_s(c) \cdot \sigma_c(c))) \cdot (1+g) + \qquad (4)\\
&\quad (\delta \cdot (\tilde{\alpha}(d) \cdot \sigma_d(c) + \tilde{\alpha}(c) \cdot \sigma_c(c)) + (1-p) \cdot \sigma_\phi(c)) \cdot (1+g) - \eta_s(c) \cdot g
\end{aligned}$$

This is because a player obtains $1+g$ points if his opponent defects (which happens with probability $\sigma_m(c)$ where $m$ is the past action observed by the opponent), and he suffers a cost of $g$ points if he cooperates. Substituting $\eta_s(c) = 1 - \eta_s(d)$ in (4) reveals that $\pi_s(\sigma)$ depends on the policy $s$ only through its dependency on $\eta_s(d)$ (the probability of defection of policy $s$). Thus we can denote $\pi_s(\sigma)$ as a function of $\eta_s(d)$: $\pi_s(\sigma) := \pi(\eta_s(d) | \sigma)$. Moreover, $\pi(\eta_s(d) | \sigma)$ is a linear function of $\eta_s(d)$. Thus, either of the following cases hold:

1. $\pi(\eta_s(d) | \sigma^*)$ is weakly increasing in $\eta_s(d)$ and thus $\eta_s(d) = 1$ is an optimal defection probability. If $\eta_\sigma(d) < 1$, then mutants who always defect, outperform the incumbents (they fare weakly better against incumbents, and strictly better against other mutants.) Thus if $\sigma$ is neutrally stable then $\eta_\sigma(d) = 1$.

2. $\pi(\eta_s(d) | \sigma^*)$ is strictly decreasing in $\eta_s(d)$ and thus $\eta_s(d) = 0$ is the unique optimal defection probability. This implies that if $\sigma$ is neutrally stable then $\eta_\sigma(d) = 0$ (otherwise, a sufficiently small group of mutants who always cooperate outperform the incumbents). If $\delta > 0$, it implies that all the incumbents must cooperate after all observed signals but then we get a contradiction to $\pi(\eta_s(d) | \sigma^*)$ being a decreasing function (because $\pi(\eta_s(d) | \sigma^*)$ is strictly increasing if the incumbents cooperate regardless of the signal.

□

## A.3 Proof of Theorem 1 (Strict Equilibrium is Strict Evolutionary Stable)

*Proof.* Assume that $a^*$ is a strict equilibrium of $G$. Let $(\xi_n)_{n \in \mathbb{N}}$ be a converging sequence of full support action perturbations (i.e., $M(\xi_n) \to_{n \to \infty} 0$.) For each $n$, let $s_n$ be the following policy:

$$s_n(\cdot)(a) = \begin{cases} \xi_n(a) & a \neq a^* \\ 1 - \sum_{a' \neq a^*} \xi_n(a') & a = a^* \end{cases}.$$

That is, the policy chooses $a^*$ with the maximal allowed probability regardless of the signal about the opponent. Let $\eta_n$ the unique outcome consistent with $s_n$:

$$\eta_n(a,a') = \begin{cases} \xi_n(a)\cdot\xi_n(a') & a,a'\neq a^* \\ \left(1-\sum_{a''\neq a^*}\xi_n(a'')\right)^2 & a=a'=a^*. \\ \left(1-\sum_{a''\neq a^*}\xi_n(a'')\right)\cdot\xi_n(a') & a=a^*\neq a' \end{cases}$$

We now show that there is $n_0\in\mathbb{N}$ such that for each $n\geq n_0$ the configuration $(s_n,\eta_n)$ is a strong evolutionary stable in the perturbed game $\Sigma(\zeta_n)$. Let $l(a^*)>0$ be the minimal loss from playing a different action against $a^*$:

$$l(a^*) = \min_{a\neq a^*}\left(\pi(a^*,a^*)-\pi(a,a^*)\right).$$

Let $g(a^*)\geq 0$ the maximal gain that can be achieved by playing a different action profile:

$$g(a^*) = \max_{(a,a')\in A\times A}\left(\pi(a,a')-\pi(a^*,a^*)\right).$$

Let $\bar{\epsilon}<$ be sufficiently small such that:

$$(1-\bar{\epsilon})\cdot 0.5\cdot l(a^*) - \frac{\bar{\epsilon}\cdot g(a^*)\cdot(k+1)}{1-2\cdot k\cdot\bar{\epsilon}} > 0. \tag{5}$$

Let $0<\epsilon<\bar{\epsilon}$. Let $\sigma'\in\Sigma(\zeta_n)$ be a mutant strategy and let $(\sigma_{s_n,\epsilon,\sigma'},\tilde{\eta})$ be a post-entry configuration (with respect to the pre-entry configuration $(s_n,\eta_n)$ .) The definition of policy $s_n$ implies that the post entry outcome is focal (the incumbents play does not depend on the observed signal and thus they play the same policy in the pre-entry and post-entry states, i.e. $\delta=0$), and thus for each actions $a,a'\in A$: $(\tilde{\eta}_n)_s(a,a') = (\eta_n)_s(a,a')$.

It remains to show that the incumbents outperform the mutants. Let $q_0$ be the probability that an incumbent plays an action different than $a^*$at each round:

$$q_0 = 1-\sum_{a''\neq a^*}\xi_n(a'').$$

Let $q_l$ be the probability that a mutant plays action $a\neq a^*$ when being matched against an incumbent. Observe that $s_n$ is the unique strategy that minimizes the probability to play $a\neq a^*$ after all observed signals. The fact that all signals are observed with positive probability (because the perturbation has full support), implies that $q_0<q_l$. Let $0\leq q_g\leq 1$ be the probability that a mutant plays action $a\neq a^*$ when being matched against a mutant.

The consistency of outcome $\tilde{\eta}$ with the strategy $\sigma_{\epsilon,\sigma'}$ implies an upper bound on $q_g$ as follows. Consider a mutant player (he) that is matched with a mutant opponent (she). The opponent can be distinguished from an incumbent opponent only to the extent that the her induced distribution of observed signals differ from the incumbents' distribution. In each observation, she plays a different action than an incumbent with probability of at most $q_g - q_0$ when being matched with a mutant partner (which happens with probability $\epsilon$), and with probability of at most $q_l - q_0$ when being matched with an incumbent partner (which happens with probability $1 - \epsilon$). When observing a match of the opponent with a mutant partner, the player may observe a different (non-incumbent) behavior also due to the action of the partner, which may play different than the incumbents with probability of at most $q_g - q_0$. Thus, the total probability to observe in any of the $k$ observations a past behavior which is different than the incumbents' behavior when being matched with a mutant opponent is at most: $k \cdot (\epsilon \cdot 2 \cdot (q_g - q_0) + (1 - \epsilon) \cdot (q_l - q_0))$.

In the remaining probability, the mutant observes behavior that is the same as the incumbents' play, and plays differently than the incumbents with probability of at most $q_l - q_0$. This yields the following upper bound on $q_g$:

$$\begin{gathered} q_g - q_0 \leq k \cdot (\epsilon \cdot 2 \cdot (q_g - q_0) + (1 - \epsilon) \cdot (q_l - q_0)) + (q_l - q_0) \Leftrightarrow \\ (q_g - q_0) \cdot (1 - k \cdot \epsilon \cdot 2) \leq (k \cdot (1 - \epsilon) \cdot (q_l - q_0)) + (q_l - q_0) \Rightarrow \\ (q_g - q_0) \cdot (1 - k \cdot \epsilon \cdot 2) \leq (k + 1) \cdot (q_l - q_0) \Leftrightarrow (q_g - q_0) \leq \tfrac{k+1}{1-2 \cdot k \cdot \epsilon} \cdot (q_l - q_0) \end{gathered}$$

For a sufficiently large $n_0$ the trembles are sufficiently small such that, for each $n \geq n_0$ a mutant player yields a loss of at least $\frac{l(a^*)}{2}$ when playing different than the incumbents against an incumbent. Thus the difference between the incumbent's payoff and the mutant's payoff is at least:

$$\pi_{s_n}\left(\sigma_{\epsilon,\sigma'}, \tilde{\eta}\right) - \pi_{\sigma'}\left(\sigma_{\epsilon,\sigma'}, \tilde{\eta}\right) \geq (1 - \epsilon) \cdot 0.5 \cdot l\left(a^*\right) \cdot (q_l - q_0) - \epsilon \cdot g\left(a^*\right) \cdot \frac{k+1}{1 - 2 \cdot k \cdot \epsilon} \cdot (q_l - q_0)$$

$$= \left((1 - \epsilon) \cdot 0.5 \cdot l\left(a^*\right) - \frac{\epsilon \cdot g\left(a^*\right) \cdot (k+1)}{1 - 2 \cdot k \cdot \epsilon}\right) \cdot (q_l - q_0) \geq$$

$$= \left((1 - \bar{\epsilon}) \cdot 0.5 \cdot l\left(a^*\right) - \frac{\bar{\epsilon} \cdot g\left(a^*\right) \cdot (k+1)}{1 - 2 \cdot k \cdot \bar{\epsilon}}\right) \cdot (q_l - q_0) > 0,$$

where the last inequality is due to (5). This implies that the mutants are outperformed, and that the configuration $(s_n, \eta_n)$ is strong evolutionary stable in $\Gamma(\zeta_n)$ for each $n \geq n_0$, which implies that $(a^*, a^*)$ is a strict limit evolutionary stable configuration. □

## A.4 Proof of Theorem 2 (Defection is the Unique Stable Outcome)

We begin by proving a simple lemma that shows that in separable Prisoner's Dilemma, the payoff of policy $s$ depends only on the frequency in which the policy defects ($\eta_s(d)$). Formally:

**Definition 13.** Given a message $m \in \{c,d\}^k$, define $d(m)$ ($c(m)$) as the number of $d$-s ($c$-s):

$$d(m) = d(a_1,...,a_k) = |\{i \leq k | a_i = d\}|, \quad c(m) = c(a_1,...,a_k) = |\{i \leq k | a_i = c\}|.$$

**Definition 14.** Given an observation structure $(p_1, 0, k)$, a message $m \in M = \phi \cup \{c,d\}^k$ and a probability $0 \leq q \leq 1$ define $\Pr(m|q)$ as the probability that the opponent observes signal $m$ conditional on the player defecting (on average) with probability $q$:

$$\Pr(m|q) = \begin{cases} 1-p_1 & m = \phi \\ \frac{k!}{(d(k))!\cdot(k-d(k))!} \cdot q^{d(k)} \cdot (1-q)^{c(k)} & \text{otherwise} \end{cases}.$$

**Lemma 3.** *Assume that $G$ is a separable Prisoner's Dilemma game, and that $p_2 = 0$. If $(\sigma, \eta)$ is a configuration, then for each $s \in C(\sigma)$:*

$$\pi_s(\sigma) = \eta_s(d) \cdot g + \sum_{m \in M} \Pr(m|\eta_s(d)) \cdot \sigma_m(c) \cdot (1+g). \tag{6}$$

*Proof.* The payoff of a player in a separable Prisoner's Dilemma is equal to $g$ times his probability of defection ($\eta_s(d)$) plus $1+g$ times the probability that the opponent cooperates ( $\sum_{m \in M} \Pr(m|\eta_s(d)) \cdot \sigma_m(c)$ .) □

As the right hand side of (6) depends only on $\eta_s(d)$, we write $\pi_s(\sigma) = \pi(\eta_s(d)|\sigma)$. We now prove Theorem 2:

*Proof.* Assume to the contrary that $(\sigma, \eta)$ is a limit weak evolutionary stable configuration and that $\eta_\sigma(d) < 1$. Let $(\zeta_n, (\sigma_n, \eta_n))_{n \in \mathbb{N}}$ be a sequence such that $(\zeta_n)_n = (\xi_n, \mathcal{S}_n, \lambda_n)_n$ is a converging sequence of perturbations, each $(\sigma_n, \eta_n)$ is a weak evolutionary stable configuration in $\Gamma(\zeta_n)$ with respect to the invasion barrier of $\bar{\epsilon}$, and $(\sigma_n, \eta_n)_n \to (\sigma, \eta)$ . Given $\xi_n(d) \leq q \leq 1 - \xi_n(c)$, let $\sigma'_{q,n} \in \Sigma(\zeta_n)$ be the strategy that assigns the maximal probability to $s \equiv q$ in $\Gamma(\zeta_n)$:

$$\sigma'_{q,n}(s) = \begin{cases} \lambda_n(s) & s \in \mathcal{S}_n \\ 1 - \sum \lambda_n(s')_{s' \in \mathcal{S}_n} & s \equiv q \\ 0 & \text{otherwise} \end{cases}.$$

Let $\bar{q}_n$ be an optimal probability of defection in the post-entry population (with mass $1-\epsilon$ to the incumbents and mass $\epsilon$ to the mutants who follow $\sigma'_{\bar{q}_n,n}$):

$$\bar{q}_n \in \text{argmax}_{q\in(\xi_n(d),1-\xi_n(c))}\left(\pi\left(q_n|\sigma_{\sigma_n,\bar{\epsilon},\sigma'_{\bar{q}_n,n}}\right)\right).$$

Let $\sigma'_n := \sigma'_{\bar{q}_n,n} \in \Sigma(\zeta_n)$. The optimality of $\bar{q}_n$ implies that the mutants weakly outperform the incumbents: $\pi_{\sigma'_n}\left(\sigma_{\sigma_n,\epsilon,\sigma'_n},\tilde{\eta}\right) \geq \pi_{\sigma_n}\left(\sigma_{\sigma_n,\epsilon,\sigma'_n},\tilde{\eta}\right)$. If $\sigma_n \neq \sigma'_n$ then this inequality contradicts $(\sigma_n,\eta_n)$ being a weakly evolutionary stable configuration, and thus contradicting the assumption that $(\sigma,\eta)$ is a limit weak evolutionary stable configuration. If $\sigma_n = \sigma'_n$, then the non-trembling incumbents' play is independent of the opponent's action, but this implies that for a sufficiently large $n$, the unique optimal defection probability against $(\sigma_n,\eta_n)$ is one, which implies that for $(\sigma,\eta)$ to be a limit weak evolutionary stable configuration, it must be that $\eta \equiv d$. □

## A.5 Proof of Theorem 3 (Stable Cooperation if $g < p_2 < 1$)

*Proof.* Let strategy $\sigma^*$ assign mass one to the following policy $s^*$:

$$s^*(m) = \begin{cases} d & m = (d,c) \\ c & \text{otherwise} \end{cases}.$$

That is, the policy induces players to cooperate in all cases except when they observed that the opponent was the sole defector in the past; in this case they defect. For each $n \in \mathbb{N}$, let $\zeta_n = \left(\xi_n \equiv 0, \mathcal{S}_n \equiv \{d\}, \lambda_n = \frac{1}{n}\right)_n$ be the perturbation in which: (1) there are no trembles when choosing actions $(\xi_n \equiv 0)$, (2) players tremble with probability $\lambda_n = \frac{1}{n}$ and choose the policy that always defects. Observe that the sequence of perturbations converge to the unperturbed game (i.e., $M(\zeta_n)_{n\to\infty} \to 0$.). □

Let $\sigma_n \in \Sigma(\zeta_n)$ the strategy that assigns maximal mass to $s^*(m)$ in the perturbed game:

$$\sigma_n(s) = \begin{cases} \frac{1}{n} & s = d \\ 1-\frac{1}{n} & s = s^* \\ 0 & \text{otherwise} \end{cases}.$$

Let $\eta_n$ be the unique consistent outcome with respect to $\sigma_n$:

$$\eta_{n,s}\left(s'\right)(d) = \begin{cases} 1 & s \equiv d \\ 0 & s = s' = s^* \\ \frac{p_2}{1+\frac{n-1}{n}p_2} & s = s^*, s' \equiv d \end{cases}, \tag{7}$$

This unique consistent outcome $\eta_n$ is derived as follows :

1. $\eta_{n,s^*}\left(s^*\right)(d) = 1$ because the policy $s^*$ never induces an observation of $(d,c)$, and thus, when players with policy $s^*$ face each other they always cooperate.

2. Let $q$ be the probability that policy $s^*$ defects when facing policy $d$; observe that $q$ must be equal to $p_2$ times the probability that policy $d$ plays $(d,c)$. This latter probability is equal to the probability of being matched with policy $s^*$ (i.e., $\frac{n-1}{n}$) times the probability that policy $s^*$ cooperates when facing policy $d$ (i.e., $1-q$). Thus $q$ is the unique solution to the equation:
$$q = p_2 \cdot \frac{n-1}{n} \cdot (1-q) \Rightarrow q = \frac{\frac{n-1}{n} \cdot p_2}{1+\frac{n-1}{n}p_2}.$$

We have to show that if $n$ is sufficiently large, then $(\sigma_n, \eta_n)$ is a strong evolutionary stable configuration in the perturbed game $\Gamma(\zeta_n)$. Observe that all action profiles are observed with positive probability when the population follows the configuration $(\sigma_n, \eta_n)$. Consider a post-entry configuration after an arbitrary group of $\epsilon$ mutants invade the population.

We first show that the post-entry outcome always remains close to the pre-entry outcome. Specifically, we show that if $\epsilon << \frac{1-p_2}{p_2}$, then the behavior of the non-trembling incumbents change by at most $\frac{p_2}{1-p_2} \cdot \epsilon$ when they face each other (it is immediate that the behavior among the incumbents when at-least one of them trembles remains the same.) Let $\alpha$ be the post-entry probability that an incumbent who follows policy $s^*$ defects when being matched with policy $s^*$. Such an incumbent defects only if he observes the opponent (and his past's partner) to be play $(d,c)$ in the past. This happens with probability $\alpha$ if the past partner was an incumbent and probability of at most 1 if the partner was mutant (and the partner was a mutant with probability $\epsilon$.) This induces the following inequality on post-entry probability of defection $\alpha$:

$$\alpha \leq p_2 \cdot ((1-\epsilon)\cdot\alpha + \epsilon) \Rightarrow \alpha \leq \frac{p_2 \cdot \epsilon}{1-(1-\epsilon)\cdot p_2} \approx \frac{p_2}{1-p_2} \cdot \epsilon,$$

where the last approximation is implied by the assumption that $\epsilon << \frac{1-p_2}{p_2}$.

Next we show that any group of mutants is outperformed. The strategy of the mutants can

differ from the incumbents strategy in one (or more) of the following ways:

1. Cooperation after observing $(d, c)$. Consider a mutant strategy in which the non-trembling mutants cooperate with positive probability after observing $(d, c)$. Observe that players observe $(d, c)$ only when being matched with the trembling policy $d$ that always defects. Thus, cooperating with positive probability yields a strict loss of $g$, when cooperating against policy $d$ without providing any gain, and the mutants are strictly outperformed.

2. Defection after observing $(c, d)$ or $(c, c)$. Consider a mutant strategy in which the non-trembling mutants defect with positive probability after observing $(c, d)$ or $(c, c)$. Note that with probability of $1 - O(\epsilon)$ such observations occur when being matched with (non-trembling) incumbents. Each such defection against the incumbents yields a gain of $g$ (the direct gain at the current interaction) and a loss of at least $1 \cdot p > g$ (the indirect loss from the fact that this behavior is observed by an incumbent in a future interaction, and it induces him to defect.) Thus, such mutants are strictly outperformed.

3. Defection after observing $(d, d)$. Consider a mutant strategy in which the non-trembling mutants defect with positive probability after observing $(d, d)$. If the mutants are sufficiently rare, most observations of $(d, d)$ are obtained when the past interaction was between the trembling policy $d$ and a non-trembling incumbent policy $s^*$,[19] and the opponent can have either of these policies with equal probability. If the opponent has policy $d$ , then defection yields a gain of $g$. If it has policy $s^*$, then defection yields a net loss of 1: a direct gain of $g$ from the current interaction, and a loss of approximately $(1 + g) \cdot p$>p>g due the probability that a future incumbent observes the current action profile of $(d, c)$. Thus, if $g < 1$ the mutants are strictly outperformed.

Similarly, one can show that mutants who combine two or more of these differences are outperformed as well. This implies that $\sigma_n(s)$ is a strong evolutionary stable configuration in the perturbed game $\Gamma(\zeta_n)$, which implies that $s^*$ is a limit strong evolutionary stable configuration.

## A.6 Proof of Theorem 4 (Robust Stable Cooperation with Public Signals)

*Proof.* Let $(\xi_n)_{n\in\mathbb{N}}$ any converging sequence of full support action perturbations (i.e., $M(\xi_n) \to_{n\to\infty} 0$.) For each $n$, let $s_n$ be the following policy (which is independent of the signal):

[19] Note, that this is true also when the share of mutants $\epsilon$ is larger than the set of the trembling policy: $\frac{1}{n} < \epsilon << 1$. Minor adaptations to the arguments following (7) show that in the unique stable outcome, the mutants probability of defection is $O\left(\frac{1}{n}\right)$, and thus observation of $(d, d)$ that involves a mutant is $\epsilon$ times less likely than observation of $(d, d)$ between the incumbents.

$$s_n(m_1,m_2)(d)=\begin{cases}1-\xi_n(c) & m_1=(d,c) \text{ or } m_2=(d,c)\\ \xi_n(d) & otherwise\end{cases}.$$

That is, the policy chooses $d$ with the maximal allowed probability of either of player was the sole defector in the past, and chooses $c$ with the maximal allowed probability otherwise. Note, that each strategy $s_n$ admits a unique outcome $\eta_n$ in which players play $(c,c)$ with probability of $1-O(M(\xi_n))$. Observe that in the limit the perturbed configurations converge to a state in which everyone cooperates with probability one:

$$s^*(d)=\lim_{n\to\infty}s_n(\cdot)(d)=\begin{cases}1 & m_1=(d,c) \text{ or } m_2=(d,c)\\ 0 & otherwise\end{cases}, \text{ and } \eta^*=\lim_{n\to\infty}\eta_n(\cdot)\equiv \text{c}.$$

We now show that there is $n_0\in\mathbb{N}$ such that for each $n\geq n_0$ the configuration $(s_n,\eta_n)$ is a strong evolutionary stable in the perturbed game $\Sigma(\zeta_n)$. Let $\epsilon<<p-g_2$. Let $\sigma'\in\Sigma(\zeta_n)$ be a mutant strategy and let $(\sigma_{s_n,\epsilon,\sigma'},\tilde{\eta})$ be a post-entry configuration (with respect to the pre-entry configuration $(s_n,\eta_n)$). Note that when two incumbents meet each other they play $(c,c)$ with probability of $1-O(M(\xi_n)+\epsilon)$. This observation can be used to show that for each mixture strategy $\sigma_{s_n,\epsilon,\sigma'}$ there exists a unique post entry consistent outcome $\tilde{\eta}$. In order to complete the proof we have to show the mutants are strictly outperformed. (for brevity, we do not present full details of the arguments and the explicit construction of $\tilde{\eta}$, which are similar to the constructions of unique post-entry configurations in Theorem 3 and related arguments in previous proofs.)

The Mutants' strategy can differ from the incumbents strategy in either of the following ways: □

1. Cooperation after observing $(d,c)$ (by either player). Consider a mutant strategy in which the non-trembling mutants cooperate with positive probability after observing $(d,c)$. With high probability (1-$O(\epsilon)$), the opponent is an incumbent, and in cooperation yields a strictly lower payoff: an incumbent is going to defect with high probability after observing $(d,c)$, and thus cooperation against him yields with high probability an immediate loss of $l$ without any indirect gain from observations by future opponents. Thus, cooperating with positive probability after observing $(d,c)$ yields a strict loss if the mutants are sufficiently rare.

2. Defection with after observing $m_1,m_2\neq(d,c)$. Consider a mutant strategy in which the non-trembling mutants defect with positive probability after observing a signal in which both $m_1,m_2\neq(d,c)$. With high probability, the opponent is an incumbent and he is going

to cooperate. In this case, defection yields an immediate gain of $g$, and an indirect loss of at least one utility point when the action profile $(d, c)$ is being observed by a future incumbent opponent (which happens with probability $p_2$.) Thus, the indirect loss is larger than the direct gain if $g < p_2$, and in this case the mutants yields a strict loss (when they are sufficiently rare.)

Similarly, one can show that mutants who combine both differences are outperformed as well. This implies that $\sigma_n(s)$ is a strong evolutionary stable configuration in the perturbed game $\Gamma(\zeta_n)$, which implies that $(s^*, \eta^*)$ is a limit strong evolutionary stable configuration.

## References

Alger, Ingela, & Weibull, Jörgen W. 2013. Homo Moralis - Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica*, **81**(6), 2269–2302.

Becker, Gary S. 1976. Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of economic Literature*, 817–826.

Berger, Ulrich, & Grüne, Ansgar. 2014. Evolutionary Stability of Indirect Reciprocity by Image Scoring.

Cressman, Ross. 1997. Local stability of smooth selection dynamics for normal form games. *Mathematical Social Sciences*, **34**(1), 1–19.

Dekel, Eddie, Ely, Jeffrey C., & Yilankaya, Okan. 2007. Evolution of preferences. *The Review of Economic Studies*, **74**(3), 685–704.

Ellison, Glenn. 1994. Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies*, **61**(3), 567–588.

Ely, Jeffrey C, & Välimäki, Juuso. 2002. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, **102**(1), 84–105.

Frank, Robert H. 1987. If homo economicus could choose his own utility function, would he want one with a conscience? *The American Economic Review*, 593–604.

Frenkel, S., Heller, Y., & Teper, R. 2014. The endowment effect as a blessing. *mimeo*.

Fujiwara-Greve, Takako, & Okuno-Fujiwara, Masahiro. 2009. Voluntarily separable repeated prisoner's dilemma. *The Review of Economic Studies*, **76**(3), 993–1021.

Guth, Werner, & Yaari, Menahem. 1992. Explaining reciprocal behavior in simple strategic games: An evolutionary approach. *In:* Witt, Ulrich (ed), *Explaining Process and Change: Approaches to Evolutionary Economics*. University of Michigan Press, Ann Arbor.

Hamilton, William D. 1964. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, **7**(1), 1–16.

Harrington Jr, Joseph E. 1995. Cooperation in a one-shot Prisoners' Dilemma. *Games and Economic Behavior*, **8**(2), 364–377.

Heifetz, Aviad, Shannon, Chris, & Spiegel, Yossi. 2007. What to Maximize If You Must. *Journal of Economic Theory*, **133**(1), 31–57.

Heller, Yuval. 2014. Stability and trembles in extensive-form games. *Games and Economic Behavior*, **84**, 132–136.

Heller, Yuval. forthcoming. Three steps ahead. *Theoretical Economics*.

Heller, Yuval, & Mohlin, Erik. 2014. Co-evolution of deception and preferences. *mimeo*.

Herold, Florian, & Kuzmics, Christoph. 2009. Evolutionary stability of discrimination under observability. *Games and Economic Behavior*, **67**(2), 542–551.

Kandori, Michihiro. 1992. Social norms and community enforcement. *The Review of Economic Studies*, **59**(1), 63–80.

Kandori, Michihiro. 2002. Introduction to repeated games with private monitoring. *Journal of Economic Theory*, **102**(1), 1–15.

Kim, Yong-Gwan, & Sobel, Joel. 1995. An evolutionary approach to pre-play communication. *Econometrica: Journal of the Econometric Society*, 1181–1193.

Kraines, David, & Kraines, Vivian. 1989. Pavlov and the prisoner's dilemma. *Theory and decision*, **26**(1), 47–79.

Leimar, Olof, & Hammerstein, Peter. 2001. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **268**(1468), 745–753.

Maynard-Smith, John. 1974. The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, **47**(1), 209–221.

Maynard-Smith, John. 1982. *Evolution and the theory of games*. Cambridge University Press.

Maynard Smith, John, & Price, George R. 1973. The logic of animal conflict. *Nature*, **246**, 15.

Nowak, Martin, & Sigmund, Karl. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, **364**, 56–58.

Nowak, Martin A, & Sigmund, Karl. 1998. Evolution of indirect reciprocity by image scoring. *Nature*, **393**(6685), 573–577.

Ohtsuki, Hisashi. 2004. Reactive strategies in indirect reciprocity. *Journal of theoretical biology*, **227**(3), 299–314.

Ok, Efe A, & Vega-Redondo, Fernando. 2001. On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory*, **97**(2), 231–254.

Okada, Akira. 1981. On stability of perfect equilibrium points. *International Journal of Game Theory*, **10**(2), 67–73.

Okuno-Fujiwara, Masahiro, & Postlewaite, Andrew. 1995. Social norms and random matching games. *Games and Economic Behavior*, **9**(1), 79–109.

Panchanathan, Karthik, & Boyd, Robert. 2003. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology*, **224**(1), 115–126.

Robson, Arthur J. 1990. Efficiency in evolutionary games: Darwin, Nash, and the secret handshake. *Journal of Theoretical Biology*, **144**(3), 379–396.

Robson, Arthur J. 1994. An "informationally robust equilibrium" for Two-Person Nonzero-Sum Games. *Games and Economic Behavior*, **7**, 233–245.

Robson, Arthur J, & Samuelson, Larry. 2010. The evolutionary foundations of preferences. *Handbook of Social Economics, Amsterdam: North Holland*.

Rosenthal, Robert W. 1979. Sequences of games with varying opponents. *Econometrica: Journal of the Econometric Society*, 1353–1366.

Sandholm, William H. 2010. Local stability under evolutionary game dynamics. *Theoretical Economics*, **5**(1), 27–50.

Sekiguchi, Tadashi. 1997. Efficiency in repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, **76**(2), 345–361.

Selten, Reinhard. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, **4**(1), 25–55.

Selten, Reinhard. 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, **84**(1), 93–101.

Selten, Reinhard. 1983. Evolutionary stability in extensive two-person games. *Mathematical Social Sciences*, **5**(3), 269–363.

Sethi, Rajiv, & Somanathan, E. 2001. Preference evolution and reciprocity. *Journal of economic theory*, **97**(2), 273–297.

Solan, Eilon, & Yariv, Leeat. 2004. Games with espionage. *Games and Economic Behavior*, **47**(1), 172–199.

Sugden, R. 1986. *The economics of rights, co-operation and welfare*. Blackwell Oxford.

Takahashi, Satoru. 2010. Community enforcement when players observe partners' past play. *Journal of Economic Theory*, **145**(1), 42–62.

Taylor, P.D., & Jonker, L.B. 1978. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, **40**(1), 145–156.

Thomas, Bernhard. 1985. On evolutionarily stable sets. *J. of Math. Biology*, **22**(1), 105–115.

Trivers, Robert L. 1971. The evolution of reciprocal altruism. *Quarterly review of biology*, 35–57.

van Damme, Eric. 1987. *Stability and Perfection of Nash Equilibria*. Springer, Berlin.

Van Veelen, Matthijs. 2009. Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *J. of Theoretical Biology*, **259**(3), 589–600.

Weibull, Jörgen W. 1995. *Evolutionary game theory*. The MIT press.