Observations on Cooperation

Heller, Yuval and Mohlin, Erik

University of Oxford

19 July 2015

Observations on Cooperation

Yuval Heller and Erik Mohlin*
Department of Economics, University of Oxford.

July 30, 2015

Abstract

We study environments in which agents are randomly matched to play a game, and before the interaction begins each agent observes a limited amount of information about the partner's aggregate behavior. We develop a novel modeling approach for such environments and apply it to study the Prisoner's Dilemma. We first show that defection is evolutionarily stable for any level of observability and behavioral noise. Next we classify the Prisoner's Dilemma into four categories of games, and we fully characterize when cooperation is evolutionarily stable in each of them.

JEL Classification: C72, C73, D83. **Keywords**: evolutionary stability, random matching, indirect reciprocity, Prisoner's Dilemma, image scoring, secret handshake.

1 Introduction

In many economic situations people are involved in short-term interactions that offer opportunities for both sides to cheat for their own gain at the expense of the others. The lack of future interactions between the agents limits the possibility to directly punish partners who act opportunistically, while the effectiveness of external enforcement is limited, e.g., due to incompleteness of contracts, non-verifiability of information, and court costs. In such situations an agent may obtain information about the partner's behavior in a sample of past interactions with other opponents, and condition his own behavior on this information. Examples of such situations include trade in the medieval world (see, e.g., Milgrom, North, and Weingast, 1990; Greif, 1993), face-to-face trade in the modern world (see, e.g., Bernstein, 1992; Dixit, 2003), and on-line interactions in Web sites such as eBay and Airbnb (see, e.g., Resnick and Zeckhauser, 2002; Jøsang, Ismail, and Boyd, 2007).

Overview of the Model Agents in a large population are randomly matched into pairs to play a symmetric one-shot game. Before playing the game, each agent privately draws a random sample consisting of a finite number of interactions between his partner and other opponents. For each such interaction he observes a

*Email: yuval.heller@economics.ox.ac.uk and erik.mohlin@economics.ox.ac.uk. A previous version of this paper was titled "Stable observable behavior". We would like to express our deep gratitude to Vince Crawford, Christoph Kuzmics, Bill Sandholm, Balázs Szentes, Rann Smorodinsky, Satoru Takahashi, Jörgen Weibull, and Peyton Young, as well as to seminar/workshop participants in Bielefeld University, Stockholm School of Economics, Israel Institute of Technology (Technion), the conference in honor of Abraham Neyman at the Hebrew University of Jerusalem, University of Pittsburgh, University of Oxford, the NBER Theory Workshop at Wisconsin-Madison, the KAEA session "Dynamic Cooperation: Theory and Evidence" at the ASSA 2015 meeting, and the Biological Basis of Preference and Strategic Behavior conference at Simon Fraser University, for many useful comments. Last but not least, we thank Renana Heller for suggesting the title.

signal that depends on the played action profile. We refer to the number of observed interactions, and the mapping from action profiles to signals as the *observation structure*.

Each agent follows a stationary *strategy*: a mapping that assigns a mixed action to each message that may be observed. The state of the population is described by a *strategy distribution*, in which different groups in the population follow different strategies. If one of these strategies is more successful than the others, then more agents start to follow it, reflecting a payoff monotonic dynamic process of cultural learning.[1]

Occasionally a small group of new agents are injected into the population, or a small group of old agents switch strategy. These agents (called, *mutants*) choose an arbitrary strategy, in a way that does not have to respect the payoff monotonic dynamics. If their new strategy is outperformed, then they abandon it. If it is more successful, then other agents start to follow it. Stability under such dynamics is captured by the static notion of evolutionary stability. A strategy distribution is *evolutionarily (neutrally) stable* (Maynard Smith and Price, 1973) if any sufficiently small group of mutants who follow a different strategy is strictly (weakly) outperformed.

Behavior is slightly perturbed by two kinds of noise. First, agents occasionally tremble when they take an action (à la Selten's 1975, 1983 notions of extensive-form perfection and limit ESS). These *action trembles* can also represent observation noise. Second, agents occasionally tremble when they revise their strategy choices, whereby they may end up following arbitrary strategies that are not necessarily payoff-maximizing. These *strategy mistakes* are similar to "normal-form" perfection à la Selten (1975) and to "crazy" agents à la Kreps, Milgrom, Roberts, and Wilson (1982). We refer to the distribution that determines the frequency of various trembles and mistakes as the *noise structure*, and we assume that it includes a positive (but possibly very small) component of strategy mistakes.

Observation Structures and Typology of PDs Our main focus in the paper is the case in which the underlying game is the *Prisoner's Dilemma* (henceforth PD). Each player decides simultaneously whether to cooperate or defect; if both players cooperate they obtain a payoff of one, if both defect they obtain a payoff of zero, and if one of the player defects, the defector gets $1+g$, while the cooperator gets $-l$ (see left side of Table 2 in Section 2). It is common to assume that mutual cooperation is the efficient outcome that maximizes the sum of payoffs, i.e., $g < l + 1$ (and games without this feature are called *non-standard PDs*).

We pay special attention to four kinds of observation structures:

1. *Observing actions*: Observing the partner's action in each sampled interaction. This is arguably the most frequently studied structure in the literature (see, e.g., Nowak and Sigmund, 1998; Milinski, Semmann, Bakker, and Krambeck, 2001; Engelmann and Fischbacher, 2009; Berger and Grüne, 2014.)

2. *Observing conflicts*: Observing in each sampled interaction whether there was mutual cooperation (i.e., no conflict; both partners are "happy") or not (i.e., partners complain about each other, but it is too costly for an outside observer to verify who actually defected). Such a structure captures the essence of feedback mechanisms used by Web sites such as eBay and Airbnb.

3. *Observing unilateral defections*: Observing whether or not the partner was the sole defector.

4. *Observing action profiles:* Observing the the full action profile in each sampled interaction.

We classify the PD games into 2 by 2 categories (see Figure 1 below):

[1] Our model also describes a biological process, in which the fitness is increasing in the game payoff.

1. Offensive/defensive PDs:[2] In an *offensive* PD there is a stronger incentive to defect against a cooperator than against a defector (i.e., $g > l$); in a *defensive* PD the opposite holds (i.e., $l > g$). If cooperation is interpreted as exerting high effort, then the defensive PD exhibits strategic complementarity; increasing one's effort from low to high is less costly if the opponent exerts high effort. As an illustration, consider a joint project of cowriting an academic paper in which each author can choose either to work hard or not. Working hard improves the probability that the paper will be of high quality, but the increment in expected quality is not worth the extra effort for an individual author. The offensive (defensive) PD describes papers that are likely to be of high quality if one of the authors (both authors) works hard (work hard) and the marginal contribution of the other (a single) hard-working author is relatively small.

2. Acute/mild PDs: Recall that the parameter g may take any value in $[0, l+1]$. We say that a PD is *acute* if g is in the upper half of this interval, i.e., if $g > \frac{l+1}{2}$, and *mild* if it's in the lower half. The threshold, $g = \frac{l+1}{2}$, is characterized by the fact that the gain from a single unilateral defection is exactly half the loss incurred by the partner who is the sole cooperator. Hence, unilateral defection is *mildly tempting* in mild PDs and *acutely tempting* in acute PDs. In order for an agent not to be tempted to defect against a cooperating partner in an acute (one-shot) PD he has to put more than half as much weight on the partner's payoff as he puts on his own payoff. Another interpretation of this threshold comes from a setup (which will be important for our results) in which an agent is deterred from unilaterally defecting because it induces future partners to unilaterally defect against the agent with some probability. Deterrence in acute PDs requires this probability of being punished to be more than 50%, while a probability of below 50% is enough for mild PDs.

Main Results Our first result (Theorem 1) shows that always defecting is evolutionarily stable for any observation structure and any noise structure. The reason is that defection is the unique best reply to itself: mutants who cooperate with positive probability against incumbents are strictly outperformed if they are sufficiently rare, and mutants who always defect against incumbents cannot identify other mutants, and therefore must also defect among themselves.[3]

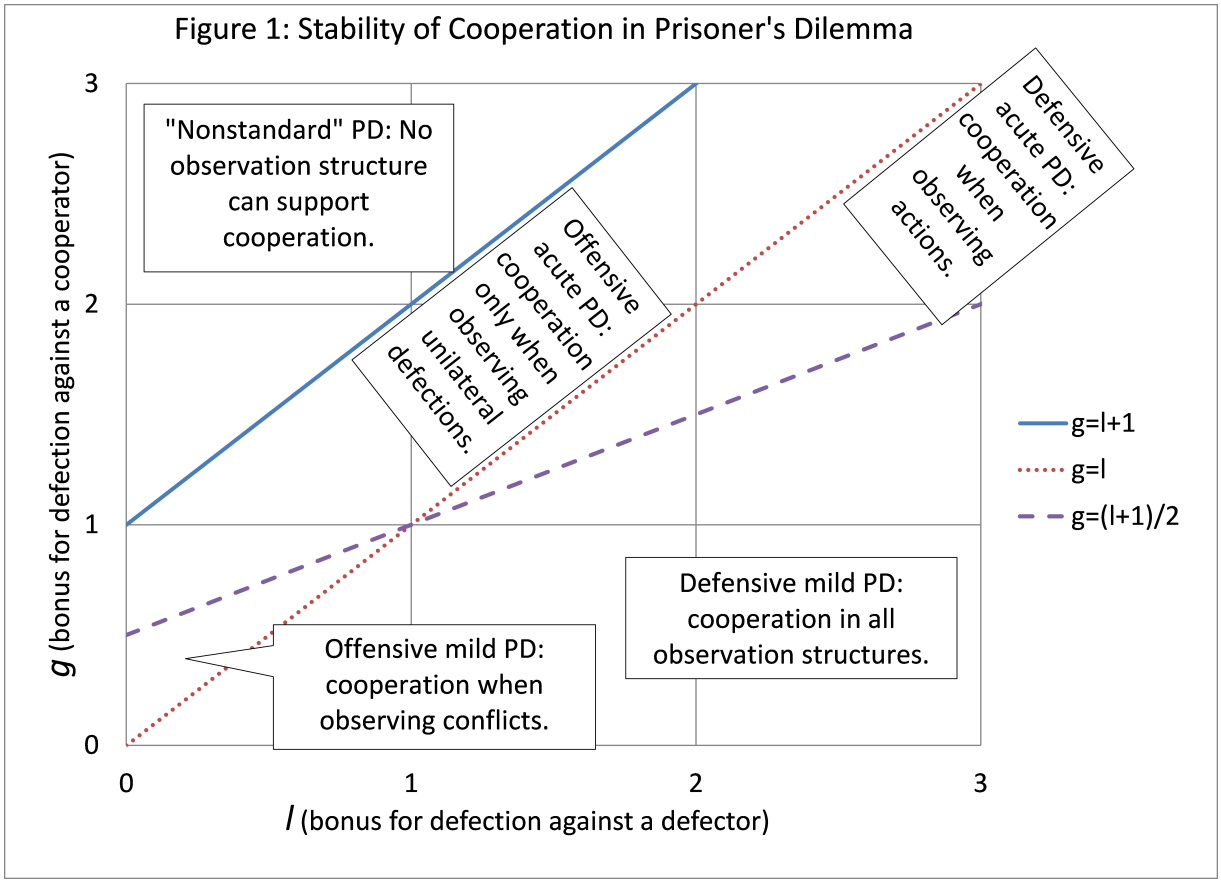
Our remaining results state under which conditions cooperation is also stable (see sketched proofs in Section 2). All of these results assume that with high probability at least two interactions are sampled . The results are summarized in Table 1 and Figure 1 below.

Table 1: Stability of Cooperation in the Prisoner's Dilemma

Category of PD	Parameters	Observation Structure (at Least 2 Sampled Interactions)			
		Actions	Conflicts	Action profiles	Unilateral Defs.
Mild & Defensive	$g < \min\left(l, \frac{l+1}{2}\right)$	Y	Y	Depends on the noise structure	Y
Acute & Defensive	$\frac{l+1}{2} < g < l$	Y	N		Y
Mild & Offensive	$l < g < \frac{l+1}{2}$	N	Y		Y
Acute & Offensive	$\max\left(l, \frac{l+1}{2}\right) < g$	N	N	N	Y

[2] This follows Dixit (2003). Takahashi (2010) calls offensive (defensive) PDs submodular (supermodular).

[3] In general games, we show that any symmetric strict equilibrium is evolutionarily stable for any observation function, provided that the noise mainly includes action trembles.



Firstly, we analyze *observation of actions*, and we show that cooperation is stable if and only if the PD is defensive. Specifically, Theorem 2 shows that defection is the unique neutrally stable outcome in offensive PDs for any noise structure and any number of observed actions. The intuition is that in offensive games, it is better to defect against partners who are likely to cooperate than against partners who are likely to defect, and this implies that mutants who always defect are more likely to induce incumbent partners to cooperate. Consequently, defecting mutants outperform incumbents who cooperate. Theorem 3 shows that cooperation is evolutionarily stable in defensive PDs for any noise distribution when the players observe at least two actions. The stability of cooperation is sustained by a population in which everyone cooperates against a partner who always cooperated, and defects against a partner who defected at least twice. In addition, with some probability (which is determined by the noise structure), agents also defect against a partner who defected at least once.[4]

Secondly, we analyze *observations of conflict*. Theorem 4 shows that cooperation is stable if and only if the PD is mild. Specifically, we show that cooperation is evolutionarily stable in any mild PD for any noise structure when agents observe at least two interactions. As in the previous result, cooperation is sustained by a population in which everyone cooperates (defects) against a partner whose sample contains no (at least two) conflicts and, in addition, agents defect with some probability against a partner who was involved in a

[4]The noise structure also determines whether each agent defects at the individual level when he observes a single defection, or whether the population is heterogeneous and includes two groups, such that agents in the first (second) group defect (cooperate) when they observe a single defection. Note that this mixed strategy is distinct from the strategy Generous Tit-For-Tat (GTFT) (Molander, 1985; Nowak and Sigmund, 1992). The latter strategy defects with positive probability even after observing two or more defections.

single conflict. In contrast, there is no noise distribution and no number of observed interactions that allow cooperation to be stable in acute PDs. The intuition is that in acute PDs, cooperation requires that any agent who is involved in a conflict be punished with a probability of at least 50%, but this implies that any conflict induce on expectation at least one more conflict, which implies that conflicts are "contagious" and induce everyone to defect.

Thirdly, we analyze *observations of unilateral defections*, and we show (Theorem 5) that when agents observe at least two interactions, cooperation is evolutionarily stable in any PD and any noise structure (and the supporting heterogeneous population is analogous to the one in the previous results).[5] This implies that if a central planner can design the observation structure, then she can pick no better mechanism than the one that only allows players to observe whether there were unilateral defections or not.

Finally, we analyze *observation of action profiles*, and we show that revealing the entire action profile may be bad for supporting cooperation. Specifically, Theorem 6 shows that cooperation is not stable with respect to any (some) noise structure in any acute (mild) PD when agents observe action profiles. The intuition is that the severity of punishments required in acute PDs implies that the conditional probability of the partner following a strategy mistake is lower when a player observes unilateral defection than when he observes a bilateral defection, which implies that unilateral defections are punished less severely than bilateral defections, which does not allow one to sustain stable cooperation. In mild PDs the same argument works if and only if the strategy mistakes lead agents to play strategies that defect with high probability.

Variants and Extensions We study four variants and extensions. First we discuss how to extend our results to a setup in which agents may choose non-stationary strategies that condition their play on their own past behavior (and we discuss the dynamic interpretation of our static model). Second, we discuss how to extend our results to the case in which both messages are observed by both players (public signals). The third extension sketches how to apply the model to study the evolution of subjective preferences that may differ from the material payoffs. Finally, we study how the invasion barriers of cooperation and defection change as the number of observations increases.

Contribution and Related Literature A substantial literature studies the possibility to sustain stable cooperation when agents in a large population are randomly matched to play the PDs (see, e.g., Nowak and Sigmund's (2005) survey on indirect reciprocity). The literature mainly studies four different setups: (1) *strangers*: players receive no information about the partner's behavior, (2) *first-order information*: a players observes past interactions that his partner was involved in (as in our model), (3) *second-order information*: a player observes what information the partner had in past interactions, and (4) *binary reputation*: a player observes a binary "label" (*e.g., good* or *bad*) about the partner, which is automatically updated according to her behavior by some external reputation mechanism.[6]

It is well known that if the population is infinite and the matching is uniform, then defection is the unique stable outcome of the PD when there are *no observations*: the *strangers*, information condition. A few papers (e.g., van Veelen, García, Rand, and Nowak, 2012; Alger and Weibull, 2013) support stable cooperation by assuming that the matching is sufficiently assortative, i.e., that cooperators are more likely to interact with

[5] We further show that in "non-standard" PDs in which $g > l + 1$ and mutual cooperation is no longer the efficient outcome, no observation structure can sustain stable cooperation.

[6] See also Rosenthal (1979) who presented an early model for random matching with observation of the partner's last action; and Wiseman and Yilankaya (2001) that show that cooperation occurs at a positive fraction of the time in a PD with pre-play communication, which can be used as "secret handshake" à la Robson (1990).

other cooperators.[7] Our paper shows that letting players observe the partner's behavior in two interactions is enough to support stable cooperation without assuming assortativity.

Kandori (1992) and Ellison (1994) (see also Deb, 2012) analyze finite populations, and show that if players are sufficiently patient, then stable cooperation can be supported by "contagious" equilibria: if one player defects at stage t, his partner defects at period $t+1$, infecting another player who defects at period $t+2$, and so on. These "contagious" equilibria have two main drawbacks: (1) a single "crazy" agent who always defects is enough to induce everyone to defect (see Ellison, 1994, p. 578), and (2) experimental evidence suggests that people typically do not follow contagious strategies (see, e.g., Duffy and Ochs, 2009). Our contribution with respect to this literature is to show that letting players observe two of the partner's past actions is enough to sustain stable cooperation, in a way that is robust to crazy agents, and that is consistent with experimental evidence on intuitive "tit-for-tat"-like behavior (see, e.g., Wedekind and Milinski, 2000; Seinen and Schram, 2006; Dal Bó and Fréchette, 2015).

In an influential paper, Nowak and Sigmund (1998) present the mechanism of *image scoring* to support cooperation when players observe the partner's past actions (first-order information). In this mechanism, each player observes several past actions of the partner, and he defects if and only if the partner's frequency of defection is above some threshold (see also the recent extension in Berger and Grüne, 2014). Experimental evidence suggests that observation of past actions substantially increases the level of cooperation (though the level of cooperation is somewhat lower than with second-order information or binary reputation mechanisms), and that many subjects seem to follow image-scoring strategies (see, e.g., Milinski, Semmann, Bakker, and Krambeck, 2001; Engelmann and Fischbacher, 2009). A few papers have raised concerns about the stability of cooperation with image-scoring mechanisms. Specifically, Leimar and Hammerstein (2001) demonstrated in simulations that cooperation is unstable, and Panchanathan and Boyd (2003) analytically studied the case in which each agent observes a single action. Our paper makes two key contributions with respect the literature on image scoring. First, Theorem 2 shows that image scoring cannot sustain stable cooperation in defensive PD games, regardless of the number of observed actions. Second, Theorem 3 presents a novel variant of image scoring, and proves that it can support cooperation in any defensive PD even when the players observe only two past actions. Our analysis shows the importance of differing between defensive and offensive PDs when modeling real-life behavior, as each kind of PD leads to qualitatively different predictions, and implies novel testable predictions for lab experiments.[8]

Takahashi (2010) studies the stability of cooperation when a player observes the *entire history* of the partner's past play. Takahashi shows that there is a sequential (strict) equilibrium that induces cooperation in any PD (if and only if the PD is defensive). Our analysis makes two related key contributions. First, we show that cooperation is stable in any defensive PD also when players observe two actions only. Second, we show that defection is the unique stable outcome in offensive PDs also when using the mild, evolutionarily motivated, solution concept of neutral stability rather than the strong notion of strict equilibrium; in particular, this implies that none of the sequential equilibria in Takahashi (2010) are neutrally stable.[9]

[7] See also Herold (2012) that studies the evolution of cooperation in a related "haystack" model in which individuals interact within separate groups, and Fujiwara-Greve and Okuno-Fujiwara (2009) that study stable cooperation in a "voluntarily separable" repeated PD, in which each player can unilaterally end and start with a randomly assigned new partner.

[8] To the best of our knowledge there is no experimental data about the influence of various values of l and g on the rate of cooperation in the PD interactions with random matching and observations of past actions. Our model predicts novel and qualitatively different comparative statics than those observed in experiments of the repeated PD (played by the same two partners); see, e.g., Blonski, Ockenfels, and Spagnolo (2011); Dal Bó and Fréchette (2011); Breitmoser (2015).

[9] In Heller (2015a) one of the authors of this paper adapts the analysis to a repeated Prisoner's Dilemma (against the same partner) with private monitoring, and shows that all the sequential equilibria in the existing equilibria are unstable in a strong sense (they are vulnerable to an arbitrarily small group of agents who always defect).

Some papers study the stability of cooperation when players also have *second-order information*; i.e., they can observe something about the past interactions of the past opponents of the current partner. The experimental evidence suggests that this second-order information helps to achieve a somewhat higher level of cooperation, as it helps to differentiate between justified and unjustified defections. However, assessing second-order information seems to carry a substantial cognitive cost for the subjects, leading many of them to look only at first-order information (see, e.g., Bolton, Katok, and Ockenfels, 2005; Gong and Yang, 2010). Our paper shows that a coarse (and hence more easily processed) form of first-order information that indicates only whether there was a unilateral defection or not, can sustain cooperation also in an offensive PD game by using simple and intuitive strategies. This novel prediction can be tested in lab experiments (to the best of our knowledge there exist no experiments that have tested it).

In a seminal book Sugden (1986) studies *binary reputation* (also called *good standing*; see related models in Kandori, 1992; Okuno-Fujiwara and Postlewaite, 1995; Ohtsuki and Iwasa, 2006). In this mechanism, all agents initially have a "good label"; an agent obtains a "bad label" by defecting against a "good" partner. The labels in this model are determined automatically by an exogenous process. Such mechanisms of binary reputations can support stable cooperation, and they seem to fit experimental behavior well when subjects observe the reputation of each partner; see, e.g., Stahl (2013). A main drawback of this approach is the requirement of having an exogenous central mechanism that manages the reputations of all players. Without such a central mechanism, it is very demanding for a player to evaluate his partner's reputation, as it depends on long histories of play of many players (because the reputation of one agent depends on the reputations of his past partners). Our main contribution is to show that the players observing the partner's behavior in two interactions is enough to support stable cooperation without requiring an external reputation mechanism.

Finally, our paper has two interesting novel insights to convey about the design of online feedback mechanisms (see, e.g., Resnick and Zeckhauser, 2002; Jøsang, Ismail, and Boyd, 2007). We first show that the plausible feedback mechanism in which players observe conflicts (without observing which side is the cause of the conflict) can yield stable cooperation iff the PD is mild. Second, we prove that observation of unilateral defections is an optimal feedback mechanism, and that it can sustain stable cooperation in any PD.

Methodological Contribution So far we have ignored a subtle aspect of the model: a strategy distribution might not uniquely determine the behavior in the population. For example, if each agent observes a single action (for sure), then any mixed action is consistent with the strategy of playing the observed action. Here we describe how we deal with this complication in the model.

An *outcome* is a mapping that describes the mixed action played by each group in the population conditional on being matched with individuals from each other group. An outcome is *consistent* with the strategy distribution if, for any two strategies in the support of the strategy distribution, it is the case that if the observations are sampled from the outcome, then the induced play coincides with the mixed actions described by the outcome. A *configuration* is a pair consisting of a strategy distribution and a consistent outcome.

Following Dekel, Ely, and Yilankaya (2007), we say that a configuration is unstable if some small invasion can move the configuration far away, either because the invading mutants outperform the incumbents, or because the entrants' presence necessarily causes a large change in aggregate behavior. Specifically, we say that a configuration is evolutionarily (neutrally) stable if after a sufficiently small group of mutants have entered the population: (1) there is a *nearby* post-entry configuration in which the incumbents play similarly to their pre-entry behavior, and (2) the mutants are strictly (weakly) outperformed in any (at least one) nearby post-entry configuration. All of our results hold with this adaptation of evolutionary stability to

configurations.

Structure Section 2 presents motivating examples and illustrates our main results. The model is presented in Section 3, and our solution concept is introduced in Section 4. Section 5 contains our main results. We discuss variants and extensions in Section 6. Section 7 concludes. The formal proofs appear in the appendix.

2 Illustrations of the Main Results

In this section we present some motivating examples, and sketch an overview of our main results.

Table 2: Matrix Payoffs of Prisoner's Dilemma (PD) Games

	c	d
c	1 1	$-l$ $1+g$
d	$1+g$ $-l$	0 0
PD $(g, l > 0\ ,\ g < l+1)$		

	c	d
c	1 1	-0.1 1.5
d	1.5 -0.1	0 0
PD1 (Offensive Mild PD)		

	c	d
c	1 1	-15 10
d	10 -15	0 0
PD2 (Defensive Acute PD)		

Prisoner's Dilemmas The left side of Table 2 presents the payoff matrix of a PD that depends on two positive parameters g and 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., an additional payoff of g) while his opponent gets $-l$. The central payoff matrix of Table 1 presents an example (PD1) of an offensive mild PD $(0.5 \cdot (l+1) > g > l)$ with $g = 0.5$ and $l = 0.1$. The right side presents an example (PD2) of a defensive acute PD $(l < g < 0.5 \cdot (l+1))$ with $g = 9$ and $l = 15$.

Defection is Evolutionarily Stable Theorem 1 shows that defection is evolutionarily stable for any small level of noise and any observation structure. The proof can be sketched as follows. The positive level of noise implies that all possible messages are observed with positive probability. Defecting with probability one regardless of the observed message is the unique strict best reply to itself. On the one hand, mutants who always defect against incumbents must also always defect among themselves. This is because such a mutant has no way of telling whether he is being matched with an incumbent or another mutant, since the partner's observed behavior is identical to that of the incumbents. On the other hand, mutants who cooperate against incumbents with an average probability of $\alpha > 0$ cooperate against other mutants with at most an additional probability of $k \cdot \alpha$. The reason is that such a mutant can cooperate against another mutant only when he observes that the other mutant cooperates in at least one of the k observed interactions. This implies that such mutants suffer a loss of $\alpha \cdot l$ from cooperating against incumbents, while their maximal gain from inducing cooperation from fellow mutants is $\epsilon \cdot (k+1) \cdot \alpha$, where $0 < \epsilon << 1$ is the fraction of mutants in the population. Thus if ϵ is sufficiently small, then the mutants are strictly outperformed.

Only Defection is Stable in Offensive PDs when Agents Observe Actions Theorem 2 shows that always defecting is the *unique* neutrally stable strategy distribution in any offensive PD for any small level

of noise, when each agent observes the partner's actions. The sketch of the proof is as follows. Assume that strategy distribution σ is neutrally stable. The payoff of a strategy in the PD can be divided into two components: (1) direct component: defecting yields an additional g points if the partner cooperates and an additional l points if the partner defects, and (2) indirect component: the strategy's average probability of defection determines the distribution of actions observed by the partners, and thereby determines the partner's probability of defecting. For each fixed average probability of defection q the fact that the PD is offensive implies that the optimal strategy among all those who defect with an average probability of q is to defect with the maximal probability against the partners who are most likely to cooperate. This implies that all agents who follow incumbent strategies are more likely to defect against partners who are more likely to cooperate. As a result, mutants who always defect outperform incumbents because they both have a strictly higher direct payoff (because defection is a dominant action) and a weakly higher indirect payoff (since incumbents are less likely to defect against them).

Specific Structure of Observation and Noise In the following sketched proofs we assume that each player observes two of his partner's interactions before playing, and that there is a single source of noise: a fraction δ of the agents in the population defect with a probability of 20% regardless of the observed message. The assumptions are made to simplify the exposition. General noise structures (which are dealt with in the formal results) yield essentially the same stable strategy distributions, except that: (1) some noise structures induce agents to mix at the individual level (rather than having two groups of agents, each following a different deterministic strategy), and (2) the probability that a player defects when he observes a single defection depends on the noise structure (it is increasing in the average probability of defection of the strategy mistakes).

Stable Cooperation in Defensive PDs when Agents Observe Actions Theorem 3 shows that cooperation is evolutionarily stable in defensive PDs when agents observe actions.

We now demonstrate how to sustain stable cooperation in PD1 (Table 2). Consider a heterogeneous distribution with two strategies in its support: 70.6% of the population defect iff they observed two defections (*Tit-For-2-Tats* strategy, abbreviated $TF2T$), and the remaining agents (29.4%) defect iff they observed at least one defection (*Tit-For-Tat* strategy, abbreviated TFT). In what follows we sketch an explanation of why this distribution is evolutionarily stable and can yield full cooperation when noise vanishes ($\delta \to 0$). We focus on outcomes in which the average probability that TFT- or $TF2T$ agents defect is $O(\delta)$. Throughout the calculations we neglect terms of $O(\delta)$ and $O(\delta^2)$ when the leading term is $O(1)$ and $O(\delta)$, respectively.

Table 3: Frequency of Defections in the Stable Population in PD1

Probability of Defection by Row str.				Calculation of Conditional Probabilities			
Frequency	29.4%	70.6%	δ	Pr (d|str.)	Pr (d,c|str.)	Pr (d,c|str.) · Pr (type)	Pr (str.|d,c)
Strategy	TFT	$TF2T$	Noise				
TFT	$2 \cdot \delta$	$0.08 \cdot \delta$	36%	δ	$2 \cdot \delta$	$0.59 \cdot \delta$	61%
$TF2T$	$O(\delta^2)$	$O(\delta^2)$	4%	$0.04 \cdot \delta$	$0.08 \cdot \delta$	$0.06 \cdot \delta$	6%
Noise	20%	20%	20%	20%	32%	$0.32 \cdot \delta$	33%

First, we calculate the defection probability of each strategy against each other strategy, as presented in the left side of Table 3. The event of a player observing a single defection is denoted by (d, c). The various probabilities are consistent with the strategies in the sense that: (1) the TFT probability of defection against

each partner is equal to the probability that a TFT-player observes the partner defecting at least once in two random interactions (i.e., it is $1-(1-\Pr(\text{d}|\text{str.}))^2$, where $\Pr(\text{d}|\text{str.})$ is the probability that the column strategy, abbreviated to str., will defect); (2) the $TF2T$ probability of defecting against a partner is equal to the probability that a $TF2T$ player observes the partner defect twice in two random interactions (i.e., it is $\Pr(\text{d}|\text{str.})^2$); and (3) noisy agents defect with a probability of 20%.

Next we have to show that both actions are best replies when agents observe a single defection. The right side of Table 3 calculates the probability that a player observes a single defection conditional on being matched with each strategy. By using Bayes' rule, we can then calculate the probability that a player is matched with each strategy conditional on the player observing a single defection. It turns out that the conditional probability of being matched with a noisy agent is ~33%. As noisy (non-noisy) agents defect with a probability of 20% ($O(\delta)$) against a non-noisy partner, it implies that the probability that a partner will defect conditional on the player observing a single defection by the partner is $33\% \cdot 20\% \approx 6.5\%$. As a result, a player's direct expected gain from defecting, conditional on having observed a single defection, is $6.5\% \cdot l + 93.5\% \cdot g = 6.5\% \cdot 15 + 93.5\% \cdot 9 = 9.4$, and since an observation of a single defection occurs with probability $0.294 \cdot 0.59 \cdot \delta + 0.706 \cdot 0.006 \cdot \delta + O(\delta^2) = 0.178 \cdot \delta$, a player's expected gain from defecting after observing single defections is $0.178 \cdot \delta \cdot 9.4$. This turns out to be equal to the indirect loss of defecting. To see this, note that the probability that a future partner is a TFT agent who observes a single defection is increased by $0.178 \cdot \delta \cdot 2 \cdot 29.4\%$ when one defects after observing single defections, and the TFT partner's defection yields a loss of 11 points, so that the total indirect cost, conditional on the player having observed a single defection, is $29.4\% \cdot 2 \cdot 16 \approx 9.4$.

The next step is to note that defection (cooperation) is the unique best reply after a player observes two defections (cooperations). This is because after a player has observed two defections (two cooperations), the conditional probability that the partner defects is higher (lower) than 6.5%, which implies that the direct gain from defecting is strictly larger (smaller) than the indirect future loss. This implies that any sufficiently small group of mutants who behave differently after observing two defections (cooperations) is strictly outperformed.

The fact that both actions are best replies after a player observes a single defection implies that TFT and $TF2T$ yield the same expected payoff. Moreover, the relative payoffs in this heterogeneous population are qualitatively similar to the payoffs in a "Hawk-Dove" game, where it is well known that the heterogeneous population is evolutionarily stable. To see why, consider a small group of invading mutants who defect after observing a single defection with an average probability of $q \neq 29.4\%$. These mutants are strictly outperformed due to the following argument. If $q > 29.4\%$ ($q < 29.4\%\%$), then the aggregate probability that a player defecting after observing a single defection, would induce a future opponent to defect is higher (lower) than $2 \cdot 29.4\%$ in the post-entry population, so that the indirect cost of defecting increases (decreases), while the direct benefit of defecting remains approximately the same. This implies that cooperating (defecting) after one has observed a single defection is the unique best reply, and the mutants who defect (cooperate) more often in these cases (relative to the incumbents) are outperformed.

Finally, mutants who defect with an average probability of 29.4% after they observe a single defection but mix at the individual level are outperformed because the supermodularity of defensive PDs implies that the payoff of a strategy as a function of its own defection probability is strictly convex (because defecting more often implies that the partners are more likely to defect against the agent, which makes defection more profitable in a defensive PD).[10]

[10] Other noise structures in which agents with noisy strategies defect after observing (c,c) and cooperate after observing (d,c)

Stable Cooperation in Mild PDs when Agents Observe Conflicts Theorem 4 shows that cooperation is evolutionarily stable when agents observe conflicts (i.e., whether or not there was mutual cooperation) iff the PD is mild. We now demonstrate how to sustain stable cooperation in the mild PD1. Consider a distribution with two strategies in its support: 77.2% of the population defect iff they observed two defections ($TF2T$), and the remaining agents (22.8%) defect iff they observed at least one defection (TFT). In what follows we sketch why this distribution is evolutionarily stable and can yield full cooperation when noise vanishes ($\delta \to 0$).

Table 4: Frequency of Defections in the Stable Population in PD1

Probability of Defection by Row st.				Calculation of Conditional Probabilities			
Freq.	22.8%	77.2%	δ	Pr (D|st.)	Pr (D,C|st.)	Pr (D,C) · Pr (type)	Pr (st.|D,C)
Strategy	TFT	$TF2T$	Noise				
TFT	$40{\cdot}\delta$	$1.4{\cdot}\delta$	71%	$20.3{\cdot}\delta$	$40.6{\cdot}\delta$	$9.2{\cdot}\delta$	85%
$TF2T$	$O\left(\delta^2\right)$	$O\left(\delta^2\right)$	21%	$0.65{\cdot}\delta$	$1.3{\cdot}\delta$	$1{\cdot}\delta$	10%
Noise	20%	20%	20%	46%	49.7%	$0.5{\cdot}\delta$	5%

First, we calculate the defection probability of every player's strategy against every other player's strategy, as presented in the left side of Table 4 (neglecting terms of $O(\delta)$ and $O\left(\delta^2\right)$ when the leading term is $O(1)$ and $O(\delta)$, respectively). The various probabilities are consistent with the strategies in the sense that: (1) the probability than a TFT agent defects against a partner is equal to the probability that the player observes at least one conflict; i.e., it is $1-(1-\Pr(D|\text{str.}))^2$, where D denotes the signal of having a conflict in a random interaction of the partner; (2) the probability that a $TF2T$ agent defects against a partner is equal to the probability that he observes conflicts in both of the partner's observed interactions; i.e., it is $\Pr(D|\text{str.})^2$; and (3) the noisy agents defect with a probability of 20%.

Next we have to show that both actions are best replies when one observes a single conflict. The right side of Table 4 calculates the probability that a player observes a single conflict (denoted by (D,C) in the table) conditional on being matched with each strategy. By using Bayes' rule, we can then calculate the probability of a player being matched with each strategy, conditional on the player observing a single conflict. This probability turns out to be 5% for the noisy agent. Thus the probability that a partner will defect conditional on the player observing a single conflict is $5\% \cdot 20\% \approx 1\%$, and the direct expected gain from defecting is $1\% \cdot l + 99\% \cdot g = 1\% \cdot 0.1 + 99\% \cdot 0.5 \approx 0.495$. This is equal to the indirect loss of defecting: $99\% \cdot 2 \cdot 22.8\% \cdot 1.1 \approx 0.495$. To see this, note that defection changes the signal from "no conflict" to "conflict" iff the partner cooperates (which happens with a probability of 99%), and each such rare signal of conflict is observed, with an average probability of $2 \cdot 22.8\%$ by a future TFT partner who then induces a loss of $l+1 = 1.1$. Finally, arguments similar to those presented above show that defection (cooperation) is the unique best reply after a player observes two (zero) conflicts, and that mixing at the individual level after the player observes a single defection yields a worse payoff.

Unstable Cooperation in Acute PDs when Agents Observe Conflicts We now sketch why one cannot sustain stable cooperation in any acute PD (i.e., PDs with $g > \frac{l+1}{2}$) when each player observes whether there was conflict in a sample of k interactions. Cooperation can be a stable outcome only if cooperating is a best reply to a cooperative partner. First consider a mutant with small mass of ϵ who

can induce the opposite case in which the payoff of an agent is a concave function of the agent's own defection probability. In such cases, the stable population is homogeneous, and all agents mix with the same probability after observing a single defection.

defects with a very small but positive average probability of z. A player's direct gain from defecting against a partner who cooperates is $z \cdot g$. A player's indirect loss is approximately $z \cdot k \cdot q \cdot (l+1)$, where q is the average probability that an incumbent defects after observing a single defection. The mutants are outperformed only if $z \cdot g < z \cdot k \cdot q \cdot (l+1)$. The acuteness of the PD implies that $0.5 \cdot (l+1) < g < k \cdot q \cdot (l+1) \Rightarrow k \cdot q > 0.5$.

Note that the mutants induce a fraction $2 \cdot k \cdot q \cdot \epsilon \cdot z$ of incumbents to defect due to observing the fraction of $\epsilon \cdot z$ conflicts that are directly induced by the mutants (recall that each defection induces a signal of conflict for both participating agents). Next, the fraction of $2 \cdot k \cdot q \cdot \epsilon \cdot z$ new conflicts will induce an additional fraction $(2 \cdot k \cdot q)^2 \cdot \epsilon \cdot z$ of defections by incumbents who observe these new conflicts. Iterating the process will show that the total number of induced conflicts is proportional to the sum of a geometric sequence with parameter $2 \cdot k \cdot q > 1$, which converges to infinity. This implies that an entry of a small group of defecting mutants is "contagious" in the sense that it induces all the incumbents to defect, which implies that cooperation cannot be a stable outcome.

Stable Cooperation in Any PD when Agents Observe Unilateral Defections The arguments for how players may support stable cooperation when they observe unilateral defections are similar to the previously sketched proofs for the case of when they observe conflicts. However, there is one key difference. In this observation structure every defection by a mutant induces a "bad" signal to at most one of the interacting agents (rather than to both of them), which implies that the geometric sequence of the total fraction of induced defections has a parameter of $k \cdot q$ (rather than $2 \cdot k \cdot q$), and thus the problem of a small group of mutants that induces the entire population to defect happens only when $g > l+1$ (rather than when $g > \frac{l+1}{2}$), which exactly characterizes "non-standard" PDs (in which, unlike the standard PDs, mutual cooperation is not the efficient outcome).

Unstable Cooperation when Agents Observe Action Profiles Finally, we sketch why cooperation is not stable in acute (mild) PDs when agents observe action profiles for all (some) of the noise structures. A population may support stable cooperation only if: (1) the incumbents on average defect with a positive probability of $q > 0$ when they observe a single unilateral defection (this is necessary for cooperation to be the best reply to a cooperator), and (2) the incumbents defect with a smaller average probability when they observe a single bilateral defection (this is necessary for cooperation to be the best reply to a defector).

On the one hand, a direct calculation of the behavior in the interactions between noisy and non-noisy agents (similar to those presented in the tables above) shows that the total frequency of unilateral defections of non-noisy agents is larger than those of noisy agents if either: (1) the PD is acute, because then the total frequency of non-noisy agents' unilateral defections is the sum of a geometric sequence with parameter $k \cdot q \geq 0.5$ as discussed above, or (2) the noisy agents defect with a probability of at least $\frac{2}{3}$ (this defection probability is so high as to imply that the non-noisy agents always defect against these noisy agents, and thus the noisy agents are never being observed to be the sole defectors).

On the other hand, mutual defections are very rare among incumbents ($O\left(\delta^2\right)$). This implies that most bilateral defections occur when at least one of the interacting agents follows a strategy mistake. This implies, that when a player observes a bilateral (unilateral) defection, the conditional probability that the partner follows a strategy mistake and is more likely to defect is approximately (less than) 50%. As a result it is more beneficial for agents to defect when they observe a bilateral defection, which contradicts requirement (2) above.

3 Model

3.1 Environment and Observation Structure

We present a reduced-form static analysis of a dynamic evolutionary process of cultural learning (or, alternatively, of a biological evolutionary process) in a large population of agents. The agents in the population are randomly matched into pairs and play a symmetric one-shot game G. Formally, let $G = (A, \pi)$ be a two-player symmetric normal-form game, where A is a finite set of actions ($|A| \geq 2$), and $\pi : A \times A \to \mathbb{R}$ is the payoff function. As is standard in the evolutionary game theory literature, we interpret the payoffs as representing "success" (or "fitness").

Let $\Delta(A)$ denote the set of mixed actions (distributions over A), and let π be extended to mixed actions in the usual way. We use the letter a (α) 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. We adopt this convention for all probability distributions throughout the paper.

Remark 1. The assumption that the game is symmetric is essentially without loss of generality (if G is played within a single population). Asymmetric games can be symmetrized by considering an extended game in which agents are randomly assigned to the different player positions with equal probability, and strategies condition on the assigned role (see, e.g., Selten, 1980).

An *observation structure* is a tuple $\Theta = (p, B, o)$, where $p \in \Delta(\mathbb{N})$ is a distribution (with a finite support) over the number of observed interactions, B is a finite set of *signals* that can be observed for each interaction, and the mapping $o : A \times A \to \Delta(B)$ describes the probability of observing each signal $b \in B$ conditional on the action profile played in this interaction (where the first action is the one played by the current partner, and the second action by her opponent).

Before playing the game, each player independently samples k independent interactions of his partner (where k is distributed according to p). Let M denote the set of all possible *messages* (profiles of signals) given observation structure Θ, i.e., $M = \left\{\cup_{k \in C(p)} B^k\right\}$, and let m denote an element of M. We let 0 be included in $\mathbb{N}$ and assume that B contains an empty message $\emptyset$ that is observed when $k = 0$. An *environment* is a pair $E = (G, \Theta)$, where G is the game and Θ is the observation structure.

We pay special attention to four kinds of observation structures:

1. *Observation actions*: Observing the partner's actions, i.e., $B = A$ and $o(a, a') = a$.
2. *Observation of action profiles:* $B = A^2$ and $o(a, a') = (a, a')$.
3. *Observation of conflicts* (in PDs): Observing whether or not there was mutual cooperation. That is, $B = \{C, D\}$, $o(c, c) = C$, and $o(a, a') = D$ for any $(a, a') \neq (c, c)$.
4. *Observation of unilateral defections* (in PDs): Observing unilateral defections of the partner. That is, $B = \{C, D\}$, $o(d, c) = D$, and $o(a, a') = C$ for any $(a, a') \neq (d, c)$.

In each of these four cases, we identify the observation structure Θ with the distribution p.

3.2 Strategies and Outcomes

A *strategy* is a mapping $s : M \to \triangle(A)$ that assigns a mixed action to each possible message. Let $s_m \in \triangle(A)$ denote the mixed action played by strategy s after observing message m. That is, for each action $a \in A$,

$s_m(a) = s(m)(a)$ is the probability that a player who follows strategy s plays action a after observing message m. We also let a denote the strategy $s \equiv a$ that plays action a regardless of the message.

Let S denote the set of all strategies, and let $\Sigma \equiv \triangle(S)$ denote the set of finite support distributions over the set of strategies. An element $\sigma \in \Sigma$ is called a *strategy distribution* (or simply *distribution*). Let $\sigma(s)$ denote the probability that strategy distribution σ assigns to strategy s. Given a strategy distribution $\sigma \in \Sigma$, let $C(\sigma)$ denote its support (i.e., the set of strategies such that $\sigma(s) > 0$). We interpret $\sigma \in \Sigma$ as representing a population in which $|C(\sigma)|$ strategies coexist, and each agent is endowed with one of these strategies according to the distribution of σ. When $|C(\sigma)| = 1$, we identify the strategy distribution with the unique strategy in its support (i.e., $\sigma \equiv s$), in line with the convention adopted above.

Remark 2. Our model focuses on stationary strategies in which the agent's behavior depends only on the message about the partner, but not on the agent's own past play or on time. We discuss how to interpret and relax this assumption in a dynamic setup in Section 6.1.

Given a finite set of strategies $\tilde{S} \subset S$, an *outcome* $\eta : \tilde{S} \times \tilde{S} \to \triangle(A)$ is a mapping that assigns to each pair of strategies $s, s' \in \tilde{S}$ a mixed action $\eta_s(s')$, which is interpreted as the mixed action played by an agent with strategy s conditional on being matched with a partner with strategy s'. Let $O_{\tilde{S}} \equiv (\triangle(A))^{(\tilde{S}\times\tilde{S})}$ denote the set of all outcomes defined over the set of strategies $\tilde{S}$. The strategy distribution and the outcome together determine the payoffs earned by each agent in the population. Outcome $\eta \in O_{\tilde{S}}$ is *pure* if there exists action $a \in A$ such that $\eta_s(s') = a$ for each $s, s' \in \tilde{S}$. We denote such a pure outcome by $\eta \equiv a$.

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

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

Let $\psi_{s,\sigma,\eta} \in \triangle(A \times A)$ be the (possibly correlated) mixed action profile that is played when an agent with strategy s is matched with a random partner sampled from σ. Formally, for each $(a, a') \in A \times A$, where a is interpreted as the action of the agent with strategy s, and a' is interpreted as the action of his partner:

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

Given a message $m_k = (b_i)_{1 \le i \le k} \in B^k$, let $\nu_{s,\sigma,\eta}(m_k)$ denote the probability that a profile of k independent observations of interactions between strategy s and a random partner is equal to m:

$$\nu_{s,\sigma,\eta}\left((b_i)_{1\le i\le k}\right) = \prod_{1<i\le k} \sum_{(a_i,a_i')\in A^2} m(a_i, a_i')(b_i) \cdot \psi_{s,\sigma,\eta}(a_i, a_i').$$

Let $\nu_{s,\sigma,\eta}(k) \in \Delta\left(B^k\right)$ be the induced distribution over messages in B^k (with $\nu_{s,\sigma,\eta}(0)(\emptyset) = 1$).

3.3 Consistent Outcomes, Configurations, and Payoffs

Fix environment (G, Θ). When individuals are drawn to play the game their actions are determined by their strategies and the messages they observe. Suppose that the observed messages are sampled from outcome η and the players play according to the strategy distribution σ. This induces a new outcome. We require

outcomes to be consistent with the strategy distribution in the sense that they generate observations that induce the current outcome to persist. Formally, given a distribution $\sigma \in \Sigma$, let $f_\sigma : O_{C(\sigma)} \to O_{C(\sigma)}$ be the mapping between outcomes that is induced by σ.

$$\left(f_\sigma(\eta)\right)_s(s')(a) = \sum_{k \in C(p)} p(k) \cdot \sum_{m_k \in M_k} \nu_{s,\sigma,\eta}(m_k) \cdot s(m_k)(a).$$

Outcome $\eta \in O_{C(\sigma)}$ is consistent with distribution σ if it is a fixed point of this mapping: $f_\sigma(\eta) \equiv \eta$. The standard Lemma 1 shows that each distribution admits a consistent outcome.

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

Proof. Observe that the space $O_{C(\sigma)}$ is a convex and compact subset of a 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 σ has a fixed point, which is a consistent outcome by definition. □

Some distributions induce multiple consistent outcomes. For example, if each player observes a single action for sure, then any outcome $\eta \in O_{\tilde{s}}$ is consistent with the distribution that assigns mass 1 to the "tit-for-tat" strategy ($\tilde{s}(a) = a$ for each $a \in A$). Due to this multiplicity, we introduce the notion of a configuration, namely, a pair consisting of a strategy distribution and a consistent outcome.

Definition 1. A *configuration* is a pair (σ, η), where $\sigma \in \Sigma$, $\eta \in O_{C(\sigma)}$, and $f_\sigma(\eta) \equiv \eta$.

Given a configuration (σ, η) and a strategy $s \in C(\sigma)$, let $\pi_s(\sigma, \eta)$ be the payoff of a player who follows strategy s in configuration (σ, η):

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

Given a distribution of strategies σ' with a weakly smaller support than σ ($C(\sigma') \subseteq C(\sigma)$), let $\pi_{\sigma'}(\sigma, \eta)$ be the payoff of a player with a strategy sampled according to σ' in configuration (σ, η):

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

Remark 3. In Heller and Mohlin (2015b) we show that all strategy distributions in an environment admit unique consistent outcomes iff the expected number of observed actions is less than one.

4 Noise and Evolutionary Stability

4.1 Noise Structures and Perturbed Environments

Our main results deal with assessing the stability of pure outcomes, such as the stability of populations of agents who always cooperate. However, the strategy of each incumbent describes his behavior also after he observes defections, which are never played on the equilibrium path. The stability analysis can therefore make sense only if one explicitly models the sources of off-path behaviors. In what follows, we define a broad notion of behavioral noise that allows for mistakes both in the choice of actions and in the choice of strategies.

A *noise structure* describes the relative frequency of each mistake that agents may commit. The first component, ξ, describes the relative frequency of choosing each pure action (i.e., each mapping from messages to pure actions) by mistake in each round. These mistakes are similar to the trembles in the definitions of extensive-form perfect equilibrium and limit ESS; see Selten, 1975, 1983. The second component, $\mathcal{S}$, describes a finite set of strategies that the agents may follow by mistake. The strategy mistakes are similar to the mistakes that are dealt with in normal-form perfection, and to the "crazy" strategies that are followed with small probability in reputation models such as Kreps, Milgrom, Roberts, and Wilson (1982).[11] The third component, λ, describes the relative frequency of each strategy mistake.

Definition 2. Let $E = (G, \Theta)$ be an environment. A *noise structure* is a tuple $\zeta = (\xi, \mathcal{S}, \lambda)$:

1. Function $\xi : A \to \mathbb{R}^+$ assigns a non-negative number to each action such that $\sum_{a \in A} \xi(a) \leq 1$, which describes the relative frequency of action trembles.

2. $\mathcal{S} \subseteq S$ is a finite set of strategy mistakes.

3. Function $\lambda : \mathcal{S} \to \mathbb{R}^+$ assigns a positive number to each strategy mistake $s \in \mathcal{S}$ s.t. $\sum_{s \in \mathcal{S}} \lambda(s) \leq 1$.

Remark 4. All our results remain the same if we require each $\xi(a)$ to be positive.

In what follows we focus on noise structures that contain a grain of full-support strategy mistakes. Specifically, we require that for each distribution of observed messages, there exist two different strategies in $\mathcal{S}$, and that at least one of them have full support (i.e., it plays all actions with positive probability).

Definition 3. Given strategy s and distribution $\mu \in \Delta(M)$, let $s_\mu = \sum_m \mu(m) \cdot s_m \in \Delta(A)$ be the distribution of actions played by an agent who follows strategy s and observes a message sampled from μ.

Definition 4. Noise structure $\zeta = (\xi, \mathcal{S}, \lambda)$ has a grain of full-support strategy mistakes if for each distribution over the set of observed messages $\mu \in \Delta(M)$, there exist strategy mistakes $s, s' \in \mathcal{S}$ such that (1) s_μ is totally mixed (i.e., $s_\mu(a) > 0$ for each $a \in A$), and (2) $s_\mu \neq s_{\mu'}$.

We interpret μ as the distribution of messages that is induced by the incumbent configuration. The first requirement (that s_μ be totally mixed) implies that any observed message might be the result of a strategy mistake. This rules out noise structures in which some messages can only be the result of action trembles (or cannot be induced at all). The second requirement is that there be two strategies in $\mathcal{S}$ that induce different plays, and thus the observed message may change the posterior probability about the partner's likely play. This rules out "degenerate" noise structures in which the entire population (including the mistake strategies) play exactly the same, and thus the observed message is completely irrelevant in assessing the likely action of the partner.

Given an environment $E = (G, \Theta)$, a noise structure $\zeta = (\xi, \mathcal{S}, \lambda)$, and a noise level $0 < \delta < 1$, we define $E(G, \Theta, \zeta, \delta)$ to be the *perturbed environment* in which agents make mistakes at an order of magnitude of δ, and the mistakes are distributed according to ζ. That is, each agent trembles and chooses action a by mistake with a probability of $\delta \cdot \xi(a)$ in each round, and a fraction of $\delta \cdot \lambda(s)$ of the population follows strategy $s \in \mathcal{S}$ by mistake.

To give a formal definition of $E(G, \Theta, \zeta, \delta)$, we need some auxiliary notation and concepts. Let $S_{\xi,\delta} \subseteq S$ be the set of strategies that assign a probability of at least $\delta \cdot \xi(a)$ to each action a after any observed message.

[11] See Abreu and Sethi (2003) for a model in which related behavioral types are evolutionarily stable.

This is the set of strategies that respect the noise structure ζ and the noise level δ. For each strategy $s \in \mathcal{S}$, let $s_{\xi,\delta} \in S_{\xi,\delta}$ be the projection of s onto $S_{\xi,\delta}$; that is, for each $a \in A$

$$s_{\xi,\delta}(a) = \min \left\{ \max \left\{ s(a), \delta \cdot \xi(a) \right\}, 1 - \sum_{a' \neq a} \delta \cdot \xi(a') \right\}.$$

Let $\mathcal{S}_{\xi,\delta}$ be the set of these projections. This is the set of strategy mistakes that are adjusted to respect the noise structure ζ and the noise level δ.

A strategy distribution $\sigma \in \Sigma$ is included in the convex set of *feasible perturbed strategy distributions* $\Sigma_{\zeta,\delta} \subseteq \Sigma$ iff: (1) the support consists entirely of strategies that respect the noise structure ζ and the noise level δ, i.e., $C(\sigma) \subseteq S_{\xi,\delta}$, and (2) all strategy mistakes that are adjusted to respect ζ and δ receive a weight that respects ζ and δ, i.e., for each $s \in \mathcal{S}_{\xi,\delta}$, $\sigma(s) \geq \delta \cdot \lambda(s)$. The perturbed environment $E(G, \Theta, \zeta, \delta)$ is defined similarly to the unperturbed environment $E(G, \Theta)$, except that the set of strategy distributions is limited to $\Sigma_{\zeta,\delta}$. Note that $E(G, \Theta, \zeta, 0)$ is the original unperturbed environment. Further note that if ζ has a grain of full-support strategy mistakes and $\delta > 0$, then any configuration (σ, η) such that $\sigma \in \Sigma_{\zeta,\delta}$ induces all messages with positive probability.

4.2 Post-Entry Focal Configuration

Our static concepts are intended to capture stable behavior in a dynamic process of cultural learning. We imagine a large population of agents. At each point in time every agent has a strategy that he currently follows. Agents regularly receive the opportunity to change their strategies. Such revisions go in the direction of the currently more successful strategies (i.e., payoff-monotonic selection dynamics). Occasionally a small group of agents, called *mutants*, switch to an arbitrary strategy, in a way that does not have to respect the payoff monotonic dynamics.

We consider incumbents distributed according to σ^* and a small group of invading mutants (with a small mass $\epsilon > 0$), who play a different distribution of strategies σ'. Following the entry, the post-entry distribution of strategies gives a weight of $1 - \epsilon$ to the incumbent strategy distribution and a weight of ϵ to the mutant strategy distribution. Following such an entry, the behavior of the population is assumed to converge to a consistent outcome of this post-entry strategy distribution. The speed at which behavior converges to a consistent outcome is assumed to be much faster than the speed at which the strategy distribution evolves in line with a payoff-monotonic learning process. Thus we can assume that the payoffs obtained in consistent outcomes are the ones that are relevant to the long-run composition of the strategy distribution.

Formally, given $0 < \epsilon < 1$ and two strategy distributions $\sigma^*, \sigma' \in \Sigma$ with relative masses of $1 - \epsilon$ and ϵ, let $\sigma_\epsilon = \sigma_{\sigma^*,\epsilon,\sigma'}$ denote the *ϵ-post-entry strategy distribution*:

$$\sigma_\epsilon(s) = (1 - \epsilon) \cdot \sigma^*(s) + \epsilon \cdot \sigma'(s) \quad \text{for each } s \in C(\sigma) \cup C(\sigma'),$$

and let an *ϵ-post-entry configuration* be any configuration consisting of the ϵ-post-entry strategy distribution and a consistent outcome: $(\sigma_\epsilon, \eta_\epsilon)$.

We say that a strategy is noisy if the strategy distribution assigns to it the minimal probability required by the noise structure. Formally, given strategy distribution $\sigma \in \Sigma_{\zeta,\delta}$ in noisy environment $E(G, p, \zeta, \delta)$, an incumbent strategy $s \in C(\sigma)$ is *noisy* if $\sigma(s) = \delta \cdot \lambda(s)$, and it is *non-noisy* otherwise, i.e., if $\sigma(s) > \delta \cdot \lambda(s)$.

We pay special attention to focal post-entry configurations in which the non-noisy incumbents play similarly to their pre-entry behavior.

Definition 5. Given a noisy environment $E(G,\Theta,\zeta,\delta)$, a strategy distribution $\sigma \in \Sigma_{\zeta,\delta}$, a configuration (σ,η), and numbers $\epsilon > 0$ and $\phi \geq 0$, we say that a post-entry configuration $(\sigma_\epsilon,\eta_\epsilon)$ is *ϕ-focal* if for any two non-noisy incumbent strategies $s, s' \in C(\sigma)$, and every action a, it holds that, $|\eta_s(s')(a) - (\eta_\epsilon)_s(s')(a)| \leq \phi$.

4.3 Evolutionary Stability

A strategy distribution is evolutionarily (neutrally) stable if following an entry of a small group of mutants: (1) there exists a post-entry focal configuration, and (2) the mutants are strictly (weakly) outperformed in any (at least one) post-entry focal configuration. Formally:

Definition 6. Fix a perturbed environment $E(G,\Theta,\zeta,\delta)$. The configuration (σ^*,η^*), where $\sigma^* \in \Sigma_{\zeta,\delta}$, is *evolutionarily stable* if for each strategy $\sigma' \neq \sigma^* \in \Sigma_{\zeta,\delta}$, and each $\bar{\phi} > 0$, there exists $\bar{\epsilon} > 0$ and $0 \leq \phi < \bar{\phi}$, such that for each $0 < \epsilon < \bar{\epsilon}$: (1) there exists a ϕ-focal ϵ-post-entry configuration; and (2) in any ϕ-focal ϵ-post-entry configuration $(\sigma_\epsilon,\eta_\epsilon)$:

$$\pi_{\sigma'}(\sigma_\epsilon,\eta_\epsilon) < \pi_{\sigma^*}(\sigma_\epsilon,\eta_\epsilon).$$

The configuration (σ^*,η^*) is *neutrally stable* if for each strategy $\sigma' \neq \sigma^* \in \Sigma_{\zeta,\delta}$, and each $\bar{\phi} > 0$, there exists $\bar{\epsilon} > 0$ and $0 \leq \phi < \bar{\phi}$, such that for each $0 < \epsilon < \bar{\epsilon}$: (1) there exists a ϕ-focal ϵ-post-entry configuration; and (2) there exists a ϕ-focal ϵ-post-entry configuration $(\sigma_\epsilon,\eta_\epsilon)$:

$$\pi_{\sigma'}(\sigma_\epsilon,\eta_\epsilon) \leq \pi_{\sigma^*}(\sigma_\epsilon,\eta_\epsilon).$$

The first condition requires that there be a post-entry configuration in which the outcome is close to the pre-entry behavior. If this condition is violated, then a small invasion can move the outcome far away, and thus the configuration is not stable. For example, consider an environment in which each agent observes a single action, and plays the observed action. Assume that initially the consistent outcome is that everyone cooperates. This configuration is unstable, because an arbitrarily small invasion of mutants who always defect would result in a post-entry strategy that has a unique outcome in which everyone defects.

The second condition requires the mutants to be outperformed in focal post-entry configurations: a strong requirement for evolutionary stability (strictly outperformed in all focal post-entry configurations), and a mild requirement for neutral stability (weakly outperformed in at least one focal post-entry configuration). The focus on focal post-entry configurations is motivated by informally considering the underlying dynamics, as Dekel, Ely, and Yilankaya (2007). Prior to the entry, the incumbent strategies have played against one another long enough to settle on the consistent outcome η^*, and it seems plausible that entry by a small group of new types will not undo this (see the dynamics presented in Section 6.1).

Note that when there are no observations ($p(0) = 1$), our definitions coincide with the classical definitions of evolutionary and neutral stability (Maynard Smith and Price, 1973).

Remark 5. A few comments are in order.

1. Our results about the stability of defection hold even if we require that all post-entry configurations be focal, or require that mutants outperform incumbents even in post-entry configurations that are not focal. In particular, always defecting is evolutionarily stable (Theorem 1) if the focality requirement is

modified in any of these ways. Moreover, one can show that no other strategy is perfectly evolutionarily stable (Theorem 2) with such an alternative definition (however, one cannot show that any other strategy is perfectly neutrally stable without our chosen definition of focality). However, the results on the stability of cooperation (Theorems 3–6) rely on only considering focal post-entry configurations to some extent, as the incumbent's strategy has two consistent outcomes, one in which almost everyone cooperates, and the other in which almost everyone defects. However, in these cases one can show that plausible dynamics like those presented in Section 6.1 would only yield the focal post-entry configuration.

2. Our results remain the same if we only allow homogeneous groups of mutants who follow a unique (non-noisy) strategy.

3. In Remark 9, we discuss the implication of requiring that all incumbent strategies outperform the mutants rather than only requiring that the incumbents outperform the mutants on average, and we explain why the alternative definition is arguably too strong.

4.4 Perfect Evolutionary Stability

A configuration is perfectly evolutionarily stable if it is the limit of evolutionarily stable configurations in a sequence of perturbed environments where the noise level converges to zero. Formally:

Definition 7. Fix environment $E = (G, \Theta)$. A sequence of strategies $(s_n)_n$ converges to strategy s if for each message $m \in M$ and each action a, the sequence of probabilities $(s_n)_m (a)$ converges to $s_m (a)$.

Definition 8. Fix environment $E = (G, \Theta)$. A sequence of configurations (σ_n, η_n) converges to a configuration (σ^*, η^*) if: for each pair of strategies $s, s' \in C(\sigma^*)$, there exist sequences of strategies $(s_n)_n$ and $(s'_n)_n$ such that: (1) $(s_n)_n \to s$ and $(s'_n)_n \to s'$, (2) $\sigma_n (s_n) \to \sigma^* (s)$ and $\sigma_n (s'_n) \to \sigma^* (s')$, and (3) $(\eta_n)_{s_n} (s'_n) \to (\eta^*)_s (s')$.

Definition 9. Configuration (σ^*, η^*) is *perfectly evolutionarily (neutrally) stable* in environment $E = (G, \Theta)$, if there exist a noise structure ζ with a grain of full-support strategy mistakes, a converging sequence of configurations $(\sigma_n, \eta_n)_n \to (\sigma^*, \eta^*)$, and a converging sequence of noise levels $(\delta_n)_n \to 0$, such that each configuration (σ_n, η_n) is evolutionarily (neutrally) stable in the perturbed environment $E(G, \Theta, \zeta, \delta_n)$. In this case we say that (σ^*, η^*) is perfectly evolutionarily (neutrally) stable with respect to noise structure ζ. If $\eta^* \equiv a^* \in A$, then we say that a^* is a *perfectly evolutionarily (neutrally) stable outcome*.

The definition of perfect evolutionary stability is analogous to Selten's (1975, 1983) notions of perfect equilibrium and limit ESS, with one difference: Selten's notions considered only action trembles, while our notion of stability deals with richer noise structures, and requires that they have a grain of full-support strategy mistakes (however, action trembles are allowed to be the most frequent kind of mistakes). In particular, when there is no observability ($p(0) = 1$), our definition of evolutionary stability coincides with Selten's definition of limit ESS.

The stability of a perfectly evolutionarily stable configuration depends on a specific noise structure. The following definition of strictly perfect evolutionary stability is more robust in the sense that it requires stability with respect to any noise structure (similar to strict perfection of Okada, 1981, and to strict limit ESS of Heller, 2015b). Formally:

Definition 10. Configuration (σ^*, η^*) is *strictly perfectly evolutionarily stable* in the environment $E = (G, \Theta)$, if for any noise structure ζ with a grain of full-support strategy mistakes, there exist a converging sequence of configurations $(\sigma_n, \eta_n)_n \to (\sigma^*, \eta^*)$, and a converging sequence of noise levels $(\delta_n)_n \to 0$, such that each configuration (σ_n, η_n) is evolutionarily (neutrally) stable in the perturbed environment $E(G, \Theta, \zeta, \delta_n)$.

Our analysis in Section 5.1 characterizes under what circumstances cooperation can be a stable outcome. We say that a pure outcome is strictly perfectly stable if for any noise structure there is a perfectly evolutionarily stable configuration (with respect to this noise structure) that induces this outcome.

Definition 11. Action a^* is a *strictly perfectly evolutionarily stable outcome* in the environment $E = (G, \Theta)$, if for any noise structure ζ with a grain of full-support strategy mistakes, there exist a converging sequence of configurations $(\sigma_n, \eta_n)_n \to (\sigma^*, \eta^*)$, and a converging sequence of noise levels $(\delta_n)_n \to 0$, such that: (1) each configuration (σ_n, η_n) is evolutionarily (neutrally) stable in the perturbed environment $E(G, \Theta, \zeta, \delta_n)$, and (2) $\eta^* \equiv a^*$.

5 Main Results

5.1 Stability in the Prisoner's Dilemma

Stability of Defection in all Environments Our first result shows that always defecting is evolutionarily stable in any PD game, for any observation function and any noise structure.[12] Recall that (d, d) represents the configuration in which everyone use the strategy of always defecting, which induces defection as its unique consistent outcome. Formally:

Theorem 1. *Let $E = (G, p)$ be an environment where G is a PD game. The configuration (d, d) is strictly perfectly evolutionarily stable.*

Observation of Actions The following two results show that under observation of actions the stability of cooperation crucially depends on whether the PD is offensive or defensive. In the former case ($g > l$) only defection is stable (Theorem 2), while in the latter case ($g < l$), cooperation is also stable (Theorem 3).

Theorem 2 shows that defection is the *unique* neutrally stable strategy distribution in any offensive PD.

Theorem 2. *Let $E = (G, p)$ be an environment with observations of actions, where G is an offensive PD. If (σ^*, η^*) is a perfectly neutrally stable configuration, then $(\sigma^*, \eta^*) = (d, d)$.*

Remark 6. Other outcomes may be neutrally stable in offensive PDs in two cases (both ruled out by our assumption that the noise structure has a grain of full-support strategy mistakes). First, if there is no noise at all, cooperation may be neutrally stable. The stability relies on the players cooperating iff the partner has never defected. However, if a player observes a defection then it is not in his interest to defect, because that will increase the probability of others defecting against him in the future. In order to avoid this problem one must make the implausible assumption that players never observe defections. Second, there might be a "degenerate" noise structure in which all players following noisy strategies defect with the same probability as the incumbents. This implies that the observed message is entirely uninformative about the partner's expected behavior. In this case a positive probability of cooperation might be supported by incumbents who tend to defect more after they observe messages with more frequent defections.

[12] When there is no noise at all, always defecting is only neutrally stable (and the game admits no evolutionarily stable strategies) because mutants who differ only in their off-the-equilibrium path behavior obtain the same payoff as the incumbents.

Remark 7. Relatively simple adaptations to the arguments of the proof of Theorem 2 show that, in the borderline case in which $g = l$, always defecting is the unique perfectly evolutionarily stable strategy. Moreover, this uniqueness result holds (for any $g \leq l$) also if we allow for noise structures without a grain of full-support strategy mistakes.

Theorem 3 shows that if players observe at least two actions, then cooperation is stable in any defensive PD and given any noise structure.[13] Formally:

Theorem 3. *Let $E = (G, p)$ be an environment with observations of actions, where G is a defensive PD ($g < l$), and $p \equiv k \geq 2$. Then cooperation is a strictly perfectly evolutionarily stable outcome.*

We sketched the proof and the construction in Section 2 for a specific noise structure. In what follows we sketch how to adapt the construction to any noise structure. Recall that in the supporting stable configuration the (non-noisy) agents cooperate (defect) when they observe zero (at least two) defections, and they defect with an average probability of q when they observe a single defection. Each noise structure induces (when the noise level converges to zero) a posterior probability of $0 < \mu < 1$ that the partner defects conditional on the player observing a single defection. Note that $\mu > 0$ due to our assumption that the noise structure also includes a grain of full-support strategy mistakes.

For each μ, there is a unique frequency $0 < q(\mu) < \frac{1}{2} \cdot \frac{l}{l+1}$ for which both actions are best replies for a player who observes that his partner has defected once. This frequency exactly balances the direct gain and the indirect loss from defecting against such a partner. The direct gain is $\mu \cdot l + (1 - \mu) \cdot g)$ and the indirect loss is $k \cdot q \cdot (1 + l)$, since each rare instance of defection is observed by on average a fraction of $k \cdot q$ partners, and each such observation induces the partner to defect with a probability of q and to yield a loss of $l + 1$. Given this q, both actions yield the same payoff when a player observes a single defection. The remaining arguments presented in Section 2 explain why the configuration is evolutionarily stable.

Remark 8. As discussed in the proof, the noise structure determines whether the stable population is heterogeneous and includes a group with mass q of TFT agents and a remaining group of $TF2T$ agents, or whether it is homogeneous and all (non-noisy) agents defect with a probability of q when they observe a single defection.

Remark 9. Following the entry of a small group of mutants who defect with a probability of $q' < q$ after they observe a single defection, the payoff of the mutants is less than the average incumbents' payoff, but it is more than the payoff of the $TF2T$ players. However, this does not influence the stability of the heterogeneous population of TFT and $TF2T$ agents in plausible smooth dynamics. The mutants have a lower payoff than the average payoff, and thus gradually disappear from the population, while the TFT ($TF2T$) becomes somewhat more (less) frequent. As soon as the mutants disappear, $TF2T$ outperforms TFT until the frequency of the TFT group returns to its original value of q.

Observation of Conflicts Next we show that under the observation of conflicts (i.e., whether or not there was mutual cooperation in each interaction), the stability of cooperation crucially depends on whether the PD is mild ($g < 0.5 \cdot (l + 1)$) or acute ($g > 0.5 \cdot (l + 1)$). Specifically, Theorem 4 shows that cooperation is (not) evolutionarily stable in any mild (acute) PD under any noise structure. The reader is referred to Section 2 for the sketched proofs of the remaining results.

[13] To simplify the notations and the formal proofs we assume in the results on the stability of cooperation that players observe a fixed number $k \geq 2$ of interactions. The results can be extended to the case of players observing a random number of interactions, which may include infrequent instances in which agents observe fewer than two actions.

Theorem 4. *Let $E = (G, p)$ be an environment with observation of conflicts, where G is a PD and $p \equiv k \geq 2$.*

1. *If G is a mild PD ($g < \frac{l+1}{2}$), then c is a strictly perfectly evolutionarily stable outcome.*
2. *If G is an acute PD ($g > \frac{l+1}{2}$), then c is not a perfectly neutrally stable outcome.*

Observation of Unilateral Defections Next we show that under the observation of unilateral defections (i.e., whether or not the partner was the sole defector), cooperation is strictly perfectly evolutionarily stable in any (standard) PD, while cooperation is unstable in non-standard PDs (in which mutual cooperation is not the efficient action profile) under any observation structure.

Theorem 5. *Let $E = (G, p)$ be an environment with observation of unilateral defections, where G is a PD.*

1. *If G is a standard PD ($g < l+1$), and $p \equiv k \geq 2$, then, cooperation is a strictly perfectly evolutionarily stable outcome.*
2. *If G is a non-standard PD ($g > l + 1$), then, c is not a perfectly neutrally stable outcome.*

Observation of Action Profiles Theorem 6 shows that under the observation of action profiles, cooperation is perfectly (but not strictly) stable in mild PDs: the stability is sensitive to the properties of the noise structure, and it holds only if the agents who follow noisy strategies defect with relatively small probability when they observe a profile of mutual cooperation. If the PD is acute, then stable cooperation cannot be supported.

Theorem 6. *Let $E = (G, p)$ be an environment with observation of action profiles, where G is a PD game and $p \equiv k \geq 2$.*

1. *If G is mild ($g < \frac{l+1}{2}$), then cooperation is a perfectly evolutionarily stable outcome, but it is not a strictly evolutionarily stable outcome.*
2. *If G is acute (i.e., $g > \frac{l+1}{2}$), then cooperation is not a perfectly neutrally stable outcome.*

5.2 Stability of Equilibria in Other Games

Our final result extends Theorem 1 (the stability of defection) to any strict Nash equilibrium a^* of any underlying game. However, the stability result for the general case holds only for some noise structures (defection in the PD is stable in all noise structures because it is a dominant action). In particular, it holds for noise structures in which the mistakes are either mostly (1) action trembles, or (2) strategy mistakes that assign high probability to playing action a^*. In either of these noise structures, players assign a high posterior probability to the event that the partner is going to play a^* regardless of the observed message, and thus playing a^* is the unique best reply. Formally:

Proposition 1. *Let $E = (G, p)$ be any environment. If (a^*, a^*) is a strict pure Nash equilibrium of G, then the configuration (a^*, a^*) is* perfectly evolutionarily stable.

Remark 10. An inefficient strict equilibrium (say, (a^*, a^*)) of a coordination game is stable only for the noise structures mentioned above. In contrast, if the noise structure mainly includes noisy strategies in which agents always play the same pure action, then a^* will not be stable. The intuition is as follows. Assume that

the incumbents play a^* (when there are no mistakes). Conditional on a player observing the partner mostly playing the efficient equilibrium action (say, a'), it is very likely that the partner is following a strategy mistake of playing a' with high probability. As a result the unique best reply given this observation is a'. However, this implies that mutants who always play a' will outperform the incumbents.

One can wonder whether Proposition 1 can be strengthened to demonstrate the stability of some non-strict equilibria of the underlying games. Example 1 suggests that this is not the case. It shows that the unique symmetric equilibrium of the underlying game, which is also an ESS and satisfies all the standard equilibrium refinements, is destabilized for any small positive level of observability.

Example 1. Consider the following Hawk-Dove game:

	d	h
d *(dove)*	1, 1	0.5, 1.5
h *(hawk)*	1.5, 0.5	0, 0

Each action is the strict best-reply to the other action, and $\alpha^* = (0.5, 0.5)$ is the unique symmetric Nash equilibrium, as well as an ESS of the underlying game. We now show why the configuration (α^*, α^*) is not neutrally stable if $p(0,0) < 1$. To simplify the argument we assume that each agent may observe only a single action, but the argument can be extended to arbitrary observation functions (and to any Hawk-Dove game). Consider a mutant strategy distribution that assigns equal weights to three strategies: (1) always play h, (2) always play d, and (3) play the opposite of the observed action, and play each action with equal probability if $\emptyset$ was observed. Intuitively, the past behavior of the mutants is informative as to the strategy they use, and this allows them to coordinate on avoiding the inefficient outcome. The mutants obtain the same payoff as incumbents when facing incumbents (because all actions yield the same payoff against α^*), but obtain a strictly higher payoff relative to the incumbents when facing other mutants. The reason is that when two mutants are matched they play the inefficient action profile (h,h) with a 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 (h,h) with a probability of $\frac{1}{4}$. This implies that the mutants outperform the incumbents in any post-entry configuration.

6 Variants and Extensions

6.1 Dynamical Interpretation and Non-stationary Strategies

Our static model raises two related questions: (1) Which plausible dynamics justify our static solution concept? (2) How restrictive is our focus on stationary strategies, infinite populations, and infinite-lived agents? In this section we sketch a dynamic model of a finite population of finite-lived patient players, and use it to interpret and justify our static model (while leaving the development of a comprehensive formal dynamic model to future research).

Fix a noisy environment $E(G, p, \zeta, \delta)$, where $E = (G, p)$ is the environment, $0 < \delta << 1$ is the noise level, and $\zeta = (\xi, \mathcal{S}, \lambda)$ is the noise structure. Let $\bar{k}$ be the largest number of observations in the support of the observation function p. Consider a population that includes a large even number $N >> 1$ of individuals, where each individual is endowed with a history and a strategy. The history is a tuple of the recent $M >> \bar{k}$ action profiles played by the agent and his mentor before him (as described below).[14] The strategy of an agent specifies the mixed action he plays as a function of his own history and the observation he has about the partner's past behavior. A stationary strategy is a strategy that depends only on the observation about

[14] The history can also include the observations about the past partners the agent had in these M interactions.

the partner (and not on the agent's own history). We allow for non-stationary strategies too. The feasible strategies are restricted by the minimal trembling probabilities determined by the noise structure. We assume that $\delta \cdot N$ of the agents (called crazy agents) follow strategies $\mathcal{S}$ (distributed according to λ).

In each round, the agents are randomly matched into pairs. Each agent obtains an observation about the partner (sampled from the partner's history according to the observation structure O), and then plays a mixed action according to his strategy. The realized action profile determines the payoff of each agent in that round. At the end of each round, each agent dies with a probability of $0 < \alpha << 1$. Each individual who dies is replaced with a new agent. A crazy agent is followed by an identical crazy agent who follows the same strategy. When a non-crazy agent dies, the new agent randomly chooses one of the incumbents as a *mentor* and copies the mentor's strategy and history. The probability of imitating a mentor is monotonically increasing in the mentor's average per-round payoff (c.f. Björnerstedt and Weibull (1996)). The interpretation is that the young agent joins the mentor as a student/apprentice for some time and learns his strategy, and the population relates the mentor's past to the likely future behavior of the apprentice.

Each configuration (σ, η) corresponds to a state of the population that consists of $|C(\sigma)|$ groups, each group includes $\sigma(s) \cdot N$ agents who follow stationary strategy s and have the history that is induced by outcome η. In addition, states of the population might also include non-stationary strategies (and in this case they will not correspond to configurations).

We are interested in characterizing the dynamically stable population states under the dynamics described above. As is standard in the evolutionary game theory literature, we explore stability by considering what happens to the population after an exogenous inflow of a small fraction of mutants who may follow arbitrary strategies. Consider any configuration that is not neutrally stable. Such a configuration cannot be dynamically stable because a small group of mutant agents can outperform the incumbents, and as a result more and more agents will start following the mutant strategy in the following generations. In an evolutionarily stable configuration each incumbent strategy earns the same expected payoff, and thus the relative frequencies of these strategies would remain constant. Moreover, since evolutionary stability implies that the mutant agents will be outperformed and thus be less likely to have followers in future generations, the mutant strategy will gradually disappear from the population.

Note that the fact that the incumbents follow stationary strategies (under the assumption that the population is described by a configuration) implies that an agent would not benefit from having a non-stationary strategy, except possibly on those very rare occasions when an agent has accumulated a history that is very different from the one induced by the outcome η.[15]

So far we have described the entry of mutants as discrete and exogenous events, in line with the literature on deterministic evolutionary dynamics. If we consider a stochastic evolutionary dynamic, by allowing for a steady but small influx of mutants, it is possible that the population will move away from evolutionarily stable configurations in the "ultra-long run" (see Samuelson, 1998). This will happen when a rare sequence of random events induces a large group of mutants and/or shifts the histories of many agents far away.

Remark 11 (Robustness to Sophisticated Agents). Consider an adaptation to the dynamics, such that in each round a small fraction of the population gets to revise their strategies. Each revising agent chooses a strategy that best-replies to the aggregate behavior of the population, where he assesses his expected stream of future payoffs according to a discount factor $\beta < 1$. If β is sufficiently close to one, then this adaptation does not affect the stability analysis. The definition of evolutionarily stable configurations implies that all

[15] Our model is not suitable for analysis of stability of configurations in which many agents follow non-stationary strategies, such as contagious equilibria (which are discussed in the Introduction).

agents already choose long-term payoff-maximizing stationary best replies, and that any other best reply is strictly outperformed when it has a sufficiently small (positive) mass in the population. If β is sufficiently close to one, then this holds also for non-stationary strategies, so sophisticated revising agents (who explicitly best-reply) will not take the population away from the evolutionarily stable configuration.

6.2 Public Messages

In the main model we assume that the message about the opponent's behavior is private. In some applications it might be more reasonable to assume that the messages 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., eBay), then the messages about the past behavior (e.g., the trader's feedback summary) are public. Another environment in which public messages are a good description is one in which a player observes the last actions that the partner played in the recent past. In such environments, the messages are essentially public because each player remembers his own recent history. In what follows we sketch how our results can be extended to the setup of public messages. To simplify the adaptation of the results, we assume that players also publicly observe a random continuous variable ("sunspot").

It is relatively straightforward to show that the stability of defection (Theorems 1–2) remains the same with public messages, and the proofs require only minor adaptations. That is, defection is stable in any public observation structure, and only defection is stable with public observation of actions. Moreover, all the results about the stability of cooperation in the various observation structures (Theorem 3–6) can be adapted to this setup as well. The population supporting stable cooperation in each of these cases consists of a single strategy according to which: (1) both players cooperate if both messages include only mutual cooperation, and (2) if at least one observed interaction includes defection (or conflict/unilateral defection in the other observation structures), then the players use the continuous public signal to coordinate their play, and both defect with a probability that is weakly increasing in the number of observed defections.

6.3 Invasion Barriers with Many Observations

Our main results show that in many cases both defection and cooperation are evolutionarily stable outcomes. In this section we discuss the robustness of these stable outcomes when observability becomes perfect, in the sense that players observe many interactions sampled from their partners' behavior. Specifically, we focus on deterministic observation functions in which agents observe k interactions sampled from the partner's behavior, and we study the limit as $k \to \infty$.

Define the *invasion barrier* of a pure outcome $\bar{\epsilon}$ to be the minimal size of a group of mutants that is required to either (1) outperform the incumbents, or (2) take the population's behavior closer to the opposite outcome, i.e., to increase the frequency of the opposite pure action above 50% in all $\bar{\epsilon}$-post-entry configurations. Let k denote the number of observations (either actions or action profiles). We say that the invasion barrier is $O\left(1/k\right)$ if there are numbers $c, \bar{k} > 0$ such that for each $k \geq \bar{k}$ the invasion barrier is smaller than c/k.

Our first observation is that the invasion barrier of defection is $O\left(1/k\right)$. The destabilizing mutants cooperate with a small probability of $0 < \theta << 1$ when they observe a partner who always defected, and cooperate for sure if they observe the partner to cooperate at least once. The direct loss from cooperating against the defecting incumbents is $\theta \cdot l$. The indirect gain from inducing cooperation between mutants is equal to $\theta \cdot k$ times the size of the mutant's group. If this size is larger than l/k, the mutants outperform the incumbent.

Intuitively, the mutants occasionally cooperate against the incumbents, and use this infrequent cooperation as a way to identify other mutants (i.e., a somewhat costly secret-handshake mechanism à la Robson, 1990).

Next, we observe that the invasion barrier of cooperation is also $O\left(1/k\right)$. For concreteness, we focus on the case of agents (privately) observing actions in defensive PDs. The stability of cooperation requires agents to defect with positive probability when they observe a single defection (otherwise mutants who defect with small probability could invade the population), which implies that they must defect for sure if the partner is observed to defect at least twice (because in this case the partner is more likely to defect against them than if only one defection has been observed, and so they cannot be indifferent between cooperation and defection). This implies that a group of mutants who always defect with a size of, say, $10/k$ will induce a post-entry population in which everyone defects with high probability (as each incumbent is likely to observe the partner to defect at least twice in the set of k observations).

It is possible to support stable cooperation with a uniform invasion barrier (which holds for all $k > \bar{k}$), in the case of public messages and public sunspots (as described in Section 6.2). This is because the public sunspots allow the players to moderate the punishment (probability of defection) after observing several defections, while with private signals there is no such mechanism to moderate these punishments.

6.4 Evolution of Subjective Preferences

In what follows we sketch how to extend the model to analyze the evolution of subjective preferences. Each subjective preference ordering is represented by a utility function on $A \times A$. A *preference-augmented configuration* is a triple consisting of a finite support distribution over utility functions, a strategy for each utility function, and a consistent outcome satisfying the requirement that each strategy be a subjective best reply (i.e, a Bayesian Nash equilibrium given the subjective preferences). The definitions of post-entry configurations, focality, and evolutionary stability can be adapted to this setup quite straightforwardly. One can then adapt Prop. 1 to this setup, and show that strict equilibrium of the underlying game is neutrally stable for any observation structure (and that the supporting distribution of preferences can assign mass one to the material preferences). This contradicts the main stylized result in the literature of the evolution of preferences that only efficient outcomes may be stable if the observation probability is sufficiently high.

The reason for this apparent contradiction is that the existing literature on the evolution of preferences (see, e.g., Güth and Yaari, 1992; Dekel, Ely, and Yilankaya, 2007; Herold and Kuzmics, 2009) assumes that each agent may directly observe the partner's preferences. In our model players observe past behavior and draw inferences about the subjective preferences (a "revealed preferences" approach). We think that our novel approach can be helpful in future research on the evolution of preferences since (1) it seems more plausible in some applications, (2) it avoids the issues of ignoring the possibility of "mimicking" mutants (see the discussion in Robson and Samuelson, 2010, Section 2.5), and (3) it avoids the crucial dependency of many results in the literature on non-generic preferences (e.g., Dekel, Ely, and Yilankaya, 2007, Prop. 2).

7 Conclusion

We study a setup in which individuals are randomly matched to play a game, and each player may observe messages about the partner's behavior. We mainly apply the model to study PDs. We show that defection is always evolutionarily stable, and we characterize which observation structures and which kinds of PDs allow cooperation to be sustained. The mechanism that supports cooperation is novel and intuitive.

Future Research We sketch three interesting directions for future research. The first direction, pursued in a companion paper by Heller and Mohlin, 2015a, studies a setup in which agents are allowed to exert effort in deception by influencing the message observed by the opponent. Second, our model assumes that players directly observe past actions of the partner. In many applications, it seems more plausible that agents observe only non-verifiable reports about the past interactions of their partner (e.g., the trader's feedback on eBay). Finally, some important interactions may be better modeled as asymmetric games between separate populations (e.g., interactions between consumers and professional sellers), and it will be interesting to extend our analysis to this setup.

A Proofs

A.1 Proof of Theorem 1 (Defection is Strictly Evolutionarily Stable)

Proof. Let $\bar{k} = \text{argmax}\{C(p)\}$ be the maximal number of observed interactions. Let $\zeta = (\xi, \mathcal{S}, \lambda)$ be any noise structure with a grain of full-support strategy mistakes. Let $\bar{\delta} > 0$ be a sufficiently small number with respect to $\frac{\min(l,g)}{(1+\max(g,l))\cdot\bar{k}}$. Let $(\delta_n)_n$ be any sequence of noise levels converging to 0 that satisfies $0 < \delta_n < \bar{\delta}$ for each n. For each n let d_n be the strategy that defects with a probability of $1 - \delta_n \cdot \xi(c)$ regardless of the observed message, and let $\sigma_n \in \Sigma_{\zeta,\delta_n}$ be such that $\sigma_n(d_{\delta,\xi}) = 1 - \sum_{s\in\mathcal{S}} \lambda(s)$. (That is, d_n is the strategy that defects with maximal probability, and σ_n is the strategy distribution that is closest to d in Σ_{ζ,δ_n}.) Let η_n be a consistent outcome of σ_n. It is immediate that $(\sigma_n, \eta_n) \to (d, d)$. Fix n. We have to show that (σ_n, η_n) is an evolutionarily stable configuration in the perturbed environment $E(G, p, \zeta, \delta_n)$.

Pick $0 < \epsilon < \bar{\epsilon}$. Let $\sigma_n \in \Sigma_{\zeta,\delta_n}$, $\sigma' \neq \sigma_n$, be a mutant strategy, and let $(\sigma_\epsilon, \eta_\epsilon)$ be an ϵ-post-entry configuration. It is immediate that $(\sigma_\epsilon, \eta_\epsilon)$ is 0-focal because the unique non-noisy strategy d_n plays the same way regardless of the observed message.

We now show that the mutants are strictly outperformed. The fact that $\sigma' \neq \sigma^*$ implies that the mutants cooperate with a higher probability than the incumbents when facing a d_n partner (because all messages are observed with positive probability, and σ_n is the unique distribution of strategies that minimizes the cooperation probability in Σ_{ζ,δ_n}). For each mutant strategy $s' \in C(\sigma')$, let $\beta_{s'}$ be an additional (average) cooperation probability of mutant s' beyond the minimal value due to trembles:

$$\beta_{s'} = \sum_{s\in C(\sigma_\epsilon)} \sigma(s) \cdot (\eta_\epsilon)_{s'}(s)(c) - \xi(c) \geq 0.$$

For each two strategies $s, s' \in C(\sigma_\epsilon)$, let $\alpha_{s,s'}$ be the difference in the cooperation probability of an agent who follows strategy s when facing a partner who follows strategy s' relative to facing an incumbent partner who follows strategy d_n, and let $\bar{\alpha}_{s'}$ be the maximum of all absolute values of $\alpha_{s,s'}$:

$$\alpha_{s,s'} = (\eta_\epsilon)_s(s')(c) - (\eta_\epsilon)_s(d_n)(c), \quad \bar{\alpha}_{s'} = \max_{s\in C(\sigma_\epsilon)} |(\alpha_{s,s'})|.$$

We now derive an upper bound for $\bar{\alpha}_{s'}$:

$$\bar{\alpha}_{s'} \leq \bar{k} \cdot (\beta_{s'} + (\epsilon + \delta_n) \cdot \bar{\alpha}_{s'}) \Rightarrow \bar{\alpha}_{s'} \leq \frac{\bar{k} \cdot \beta_{s'}}{1 - \bar{k} \cdot (\epsilon + \delta_n)}.$$

To see why this is the case, note that the LHS is the maximal probability that an agent plays differently when

facing a mutant s'-partner, than when facing a d_n-partner. This is bounded by the probability that the agent observes the mutant s'-partner (or any of his past opponents) play differently from what play looks like in interactions involving a d_n-partner, in any of the $\bar{k}$ observed interactions. In each such observed interaction, the partner plays differently from a d_n-agent with a probability of $\beta_{s'}$, and the partner's opponent in that interaction plays differently only if she follows either a noisy strategy or a mutant strategy (which happens with a probability of $\epsilon + \delta$), and in this case she plays differently with a probability of at most $\bar{\alpha}_{s'}$.

The s'-mutants suffer a direct loss of $\beta_{s'} \cdot \min(l, g)$ from their higher cooperation probability (relative to the d_n-agents). Their indirect gain (from inducing partners to cooperate more often against them) is at most $(\epsilon + \delta_n) \cdot \max_{s \in C(\sigma_\epsilon)} (\alpha_{s,s'}) \cdot (1 + \max(g, l)) \leq (\epsilon + \delta_n) \cdot \bar{\alpha}_{s'} \cdot (1 + \max(g, l))$. Thus the loss outweighs the gain if:

$$(\epsilon + \delta_n) \cdot \bar{\alpha}_{s'} \cdot (1 + \max(g, l)) \leq \frac{(\epsilon + \delta_n) \cdot (1 + \max(g, l)) \cdot \bar{k} \cdot \beta_{s'}}{1 - \bar{k} \cdot (\epsilon + \delta_n)} \leq \beta_{s'} \cdot \min(l, g),$$

which holds for our choice of $\bar{\epsilon}, \bar{\delta}$ as sufficiently small. □

A.2 Proof of Theorem 2 (Only Defection is Stable in Offensive PDs)

Proof. Let $\zeta = (\xi, \mathcal{S}, \lambda)$ be a noise structure with a grain of full-support strategy mistakes. Let $(\sigma_n, \eta_n)_n \to (\sigma^*, \eta^*)$ be a converging sequence of configurations, and let $(\delta_n)_n \to 0$ be a converging sequence of noise levels, such that each configuration (σ_n, η_n) is neutrally stable in the perturbed environment $E(G, p, \zeta, \delta_n)$. That is, we assume that (σ^*, η^*) is a perfectly neutrally stable configuration. In order to obtain a contradiction assume that $\sigma^* \neq d$.

Recall that any message $m \in M$ is observed with positive probability due to the noise structure. Given configuration (σ_n, η_n), message $m \in M$, and strategy $s \in C(\sigma_n)$, let $q_m(s)$ denote the expected probability that a randomly drawn partner of a player defects, conditional on the player following strategy s and observing message m about the partner.

We say that a strategy is defector-favoring if the strategy defects against partners who are likely to cooperate, and cooperates against partners who are likely to defect. Specifically, a strategy is defector-favoring if there is some threshold such that the strategy cooperates (defects) when the partner's conditional probability of defecting is above (below) this threshold. Formally:

Definition 12. Strategy $s \in C(\sigma_n)$ is *defector-favoring* given configuration (σ_n, η_n) if there is some $0 \leq \bar{q} \leq 1$ such that, for each $m, m' \in M$, $q_m(s) > \bar{q} \Rightarrow s_m(d) = 0$, and $q_m(s) < \bar{q} \Rightarrow s_m(d) = 1$.

The rest of the proof consists of the following four steps.

1. First we show that all non-noisy strategies in σ_n are defector-favoring. Assume to the contrary that there is a non-noisy strategy $s \in C(\sigma_n)$ that is not defector-favoring. Let σ' be a mutant strategy distribution that is exactly like the incumbent strategy distribution σ_n except that a positive fraction of strategy s is replaced by a strategy s' that has the same average defection probability as s in a focal ϵ-post-entry population but is defector-favoring (where $0 < \epsilon << 1$ is taken to be sufficiently small). The fact that both strategies defect with the same average probability implies that they induce the same behavior from the partners (since these partners observe identical distributions of messages when facing s and when facing s'), hence $q_m(s) = q_m(s')$. Strategy s' defects more often against partners who are more likely to cooperate relative to strategy s. Since the PD is offensive this implies that strategy s'

strictly outperforms strategy s, which implies that the mutant distribution σ' strictly outperforms the incumbent distribution σ.

2. Second we show that all the non-noisy strategies defect with the same average probability in (σ^*, η^*). Assume to the contrary that there are non-noisy strategies $s, s' \in C(\sigma^*)$ such that $\eta_{s,\sigma^*}(d) > \eta_{s',\sigma^*}(d)$. Note that agents who follow strategy s have strictly higher payoff than agents who follow s' when being matched with non-noisy agents. This is because strategy s yields: (1) a strictly higher direct payoff due to playing more often the dominant action d, and (2) a weakly higher payoff against non-noisy agents, because the fact that it defects more often and all non-noisy agents follow defector-favoring strategies implies that non-noisy partners defect with a weakly smaller probability when being matched with agents who follow strategy s (relative to s'). This implies that for a sufficiently small noise level, the followers of s would have a strictly higher payoff than the followers of s', which contradicts (σ^*, η^*) being perfectly neutrally stable (because a sufficiently small group of mutants similar to the incumbents, except that the strategy s' is replaced by s, would outperform the incumbents in any nearby focal post-entry configuration).

3. Next we show that for any non-noisy player it is the case that the probability that the partner defects conditional on the player observing a message that only includes defections (denoted by message $\overrightarrow{d}$) is weakly larger than the probability that the partner defects conditional on the player observing a message of the same length that also includes cooperation (i.e., $q_m(s) < q_{\vec{d}}(s)$ for any non-noisy strategy s and any message $m \neq \overrightarrow{d}$ with the same length as $\overrightarrow{d}$). To see why this is the case, note that the fact that the noise structure has a grain of full-support strategy mistakes implies that not all noisy strategies have the same defection probabilities, and thus the signal about the partner yields some information about the partner's probability of defecting. The previous step shows that all non-noisy agents defect with the same probability when the noise level converges to zero, which implies that if the noise level is positive but very small, then they induce almost the same signal distribution, and thus they induce almost the same behavior from all partners. Combining this fact with the fact that not all strategies have the same defection probability, implies that if a player observes a message that only includes defections, then the partner is more likely to have a higher average defection probability when being matched with any non-noisy agent (i.e., $q_m(s) < q_{\vec{d}}(s)$ for any non-noisy strategy s).

4. Thus, any non-noisy agent (who follows a defector-favoring strategy due to the first step) would defect with a weakly higher probability after observing signal $\overrightarrow{d}$. This implies that if the noise level is sufficiently small, then a mutant distribution that assigns maximal mass to the strategy that defects with the highest probability outperforms the incumbents. The mutants achieve a direct higher payoff by defecting more often, as well as a weakly higher indirect gain by inducing the incumbents to cooperate more often.

□

A.3 Proof of Theorem 3 (Stable Cooperation in Defensive PDs)

Proof. Let TFT ($TF2T$) be the strategy that defects iff the partner is observed to defect at least once (twice). Let TFT_q be the strategy that defects with a probability of q (to be defined later) iff the partner is observed to defect once, and defects for sure if he is observed to defect twice or more. Let $\zeta = (\xi, \mathcal{S}, \lambda)$ be

a noise structure with a grain of full-support strategy mistakes. Let σ be the strategy that assigns mass q (defined below) to TFT and mass $1-q$ to $TF2T$. Let σ' be the strategy that assigns mass one to TFT_q. Let $\eta = \eta' \equiv c$. Let $\bar{\delta}$ be a number that is sufficiently small relative to $\frac{1}{k\cdot(l+1)}$. For each $n \geq 1$, let $\delta_n = \bar{\delta}/n$. Let σ_n (σ'_n) be the closest strategy to σ (σ') in Σ_{ζ,δ_n}, i.e.,

$$\sigma_n(TFT_\xi) \quad = \quad \bar{\lambda}_n \cdot q, \;\; \sigma_n(TF2T_\xi) = \bar{\lambda}_n \cdot (1-q), \;\; \sigma_n(TFT_{q,\xi}) = \bar{\lambda}_n \text{ where } \bar{\lambda}_n = 1 - \delta_n \cdot \sum_{s\in\mathcal{S}} \lambda(s),$$

where strategy TFT_ξ ($TF2T_\xi$, $TFT_{q,\xi}$) is the same as TFT ($TF2T$, TFT_q), except that the probability of cooperation (defection) after each observed signal is adapted to lie on the closest boundary of the interval $[\delta_n \cdot \xi(c), 1-\delta_n \cdot \xi(d)]$ ($[\delta_n \cdot \xi(d), 1-\delta_n \cdot \xi(c)]$).

For each $s \in C(\sigma_n)$ or $s \in C(\sigma'_n)$, let $Pr(d, \overrightarrow{c}|s)$ ($Pr(d,d,...|s)$) denote the probability of observing exactly one defection (at least two defections) conditional on the partner following strategy s. Let $Pr(d, \overrightarrow{c})$ and ($Pr(d,d,...)$ be the corresponding unconditional probabilities in configurations (σ_n, η_n) and (σ'_n, η'_n), respectively. When we calculate each of these probabilities we will rely on the fact that $\delta_n << 1$. Thus we will neglect terms of $O(\delta_n)$ ($O(\delta_n^2)$) when the leading term is $O(1)$ ($O(\delta_n)$).

We will assume that the sequences of outcomes η_n and η'_n are such that $\eta_n, \eta'_n \to \eta \equiv c$ as $n \to \infty$, and then show that such η_n and η'_n are indeed consistent with σ_n and σ'_n, respectively. Note that $\eta_n, \eta'_n \to c$ implies that $\left((\eta_n)_{s,\sigma_n}(d)\right) = O(\delta_n)$ and $\left((\eta'_n)_{s,\sigma_n}(d)\right) = O(\delta_n)$ for all $s \in C(\sigma)$ and $s \in C(\sigma')$, respectively. Thus our calculations will rely on the fact that agents are very likely to observe the message $\vec{c}$ (which consists of k cooperations) from a random opponent; formally, $Pr(\overrightarrow{c}) = (1-O(\delta_n))^k = 1 - O(\delta_n)$.

The conditional probabilities for a noisy strategy $s \in \mathcal{S}$ are (with an analogous formula in η')

$$Pr(d, \overrightarrow{c}|s) = k \cdot \eta_s(\overrightarrow{c})(d) \cdot (\eta_s(\overrightarrow{c})(c))^{k-1} + O(\delta_n),$$

$$Pr(d,d,...|s) = 1 - Pr(d, \overrightarrow{c}|s) - (\eta_s(\overrightarrow{c})(c))^k + O(\delta_n).$$

To simplify the exposition (with slight abuse of notation) we let $TFT_{q,\xi}$ denote the strategy distribution that puts probability q on TFT_ξ and probability $1-q$ on $TF2T_\xi$ in configuration (σ_n, η_n). Given message m, let $Pr(m|TFT_{q,\xi})$ denote the probability of observing message m, conditional on the partner following $TFT_{q,\xi}$ in configuration (σ'_n, η'_n), or following the mix of TFT_ξ (with a probability of q) and $TF2T_\xi$ (with a probability of $1-q$) in configuration (σ_n, η_n). Thus in both configurations (σ_n, η_n) and (σ'_n, η'_n) we have:

$$Pr(m|TFT_{q,\xi}) = q \cdot Pr(m|TFT_\xi) + (1-q) \cdot Pr(m|TF2T_\xi).$$

Note that the non-noisy strategies very rarely defect twice or more in k interactions: $Pr(d,d,...|TFT_{q,\xi}) = O(\delta_n^2)$. Next we calculate the probability of a non-noisy incumbent generating a message that contains a single defection. This happens if either (1) one of the k partners is observed to defect twice, or (2) with a probability of q one of the k partners is observed to defect once, or (3) due to a tremble:

$$Pr(d, \overrightarrow{c}|TFT_{q,\xi}) = k\cdot\delta_n \cdot \sum_{s\in\mathcal{S}} \lambda(s) \cdot (Pr(d,d,...|s) + q \cdot Pr(d, \overrightarrow{c}|s)) + k\cdot q \cdot Pr(d, \overrightarrow{c}|TFT_q) + \delta_n \cdot k \cdot \xi(d) + O(\delta_n^2).$$

Solving this yields (neglecting $O\left(\delta_n^2\right)$)

$$Pr\left(d,\overrightarrow{c}|\mathrm{TFT}_{q,\xi}\right)=\frac{k\cdot\delta_n\cdot\sum_{s\in\mathcal{S}}\lambda\left(s\right)\cdot\left(Pr\left(d,d,...|s\right)+q\cdot Pr\left(d,\overrightarrow{c}|s\right)\right)+k\cdot\delta_n\cdot\xi\left(d\right)}{1-k\cdot q},$$

which is well defined and $O\left(\delta_n\right)$ as long as $q<1/k$. We can now calculate the unconditional probabilities:

$$Pr\left(d,\overrightarrow{c}\right)=\delta_n\cdot\sum_{s\in\mathcal{S}}\lambda\left(s\right)\cdot Pr\left(d,\overrightarrow{c}|s\right)+Pr\left(d,\overrightarrow{c}|\mathrm{TFT}_{q,\xi}\right)+O\left(\delta_n^2\right),$$

$$Pr\left(d,d,...\right)=\delta_n\cdot\sum_{s\in\mathcal{S}}\lambda\left(s\right)\cdot Pr\left(d,d,...|s\right)+O\left(\delta_n^2\right).$$

By using Bayes' rule we can calculate the conditional probability that the partner uses strategy $s\in C\left(\sigma_n\right)$ as a function of the observed message:

$$Pr\left(s|d,\overrightarrow{c}\right)=\frac{\sigma_n\left(s\right)\cdot Pr\left(d,\overrightarrow{c}|s\right)}{Pr\left(d,\overrightarrow{c}\right)},\quad Pr\left(s|d,d,...\right)=\frac{\sigma_n\left(s\right)\cdot Pr\left(d,d,...|s\right)}{Pr\left(d,d,...\right)}.$$

For a sufficiently large n the conditional probability that the partner follows a noisy strategy is higher the more defections there are in the observed message:

$$O\left(\delta_n\right)=\sum_{s\in\mathcal{S}}Pr\left(s|\overrightarrow{c}\right)<\sum_{s\in\mathcal{S}}Pr\left(s|d,\overrightarrow{c}\right)<\sum_{s\in\mathcal{S}}Pr\left(s|d,d,...\right).$$

To see that this is the case, note that $Pr\left(d|\,TFT_{q,\xi}\right)=O\left(\delta_n\right)$, while $Pr\left(d|s\right)=\eta_s\left(\overrightarrow{c}\right)\left(d\right)+O\left(\delta_n\right)$, for any noisy strategy, and because of the grain of full-support strategy mistakes there is at least one noisy strategy s such that $\eta_s\left(\overrightarrow{c}\right)\left(d\right)>0$.

Given a message m let $Pr\left(TFT_{q,\xi}|d,\overrightarrow{c}\right)$ in the configuration $\left(\sigma_n,\eta_n\right)$ denote the conditional probability that the partner follows either TFT_ξ or $TF2T_\xi$ (and denote the conditional probability that the partner follows $TFT_{\xi,q}$ in the configuration $\left(\sigma_n',\eta_n'\right)$). The calculations above show that we have $Pr\left(TFT_{q,\xi}|d,\overrightarrow{c}\right)=O\left(\delta_n\right)$, which implies that $\lim_{\delta_n\to0}\left(\sum_{s\text{ is noisy}}Pr\left(s|d,\overrightarrow{c}\right)\right)>0$. Let μ be the probability that a random partner defects conditional on a player observing message $\left(d,\overrightarrow{c}\right)$ about the partner, and conditional on the partner observing the message $\overrightarrow{c}$:

$$\mu=\sum_{s\in\mathcal{S}}Pr\left(s|d,\overrightarrow{c}\right)\cdot s_{\overrightarrow{c}}\left(d\right)+O\left(\delta_n\right). \tag{1}$$

Note that $O\left(\delta_n\right)<\mu$ because $\lim_{\delta_n\to0}\left(\sum_{s\text{ is noisy and }s_{\overrightarrow{c}}(d)>0}Pr\left(s|d,\overrightarrow{c}\right)\right)>0$. Eq. (1) defines μ as a strictly decreasing function of q. To see this, note that the term $s_{\overrightarrow{c}}\left(d\right)$ does not depend on q, and in $Pr\left(s|d,\overrightarrow{c}\right)=\frac{\sigma_n(s)\cdot Pr\left(d,\overrightarrow{c}|s\right)}{Pr\left(d,\overrightarrow{c}\right)}$ the terms $\sigma_n\left(s\right)$ and $Pr\left(d,\overrightarrow{c}|s\right)$ do not depend on q, whereas the term $Pr\left(d,\overrightarrow{c}\right)$ is increasing in q.

Next we calculate the value of q (fraction q of TFT agents in $\left(\sigma_n,\eta_n\right)$ or mixture probability q in $\left(\sigma_n',\eta_n'\right)$) that balances the payoff of both actions after a player observes a single defection (neglecting terms of $O\left(\delta_n\right)$). The LHS of the following equation represents the player's direct gain from defecting in the rare cases when he observes a single defection, while the RHS represents the player's indirect loss induced by partners who

defect as a result of observing these defections:

$$\mu \cdot l + (1-\mu) \cdot g = k \cdot q \cdot (l+1) \Rightarrow q = \frac{\mu \cdot l + (1-\mu) \cdot g}{k \cdot (l+1)}. \tag{2}$$

Note that Eq. 2 defines q as a strictly increasing function of μ. This implies that there are unique values of q and μ, satisfying $\frac{g}{k\cdot(l+1)} < q < \frac{l}{k\cdot(l+1)} < \frac{1}{k}$ and $0 < \mu < 1$, which jointly solve Eqs. 1 and 2. By standard continuity arguments, for any n, there exists a frequency $q_n = q + O(\delta_n)$ that balances the payoff of both actions after a player observes a single defection given the noisy distribution of strategies σ_n.

Observe that defection is the unique best reply when a player observes at least two defections. The direct gain from defecting is larger than the LHS of Eq. 2, and the indirect loss is still given by the RHS of Eq. (2). The reason that the direct gain is larger is that non-noisy partners almost never defect twice or more (the probability is $O\left(\delta_n^2\right)$), and thus the partner is most likely to follow a noisy-strategy with a defection probability that is higher than μ (since μ also gives weight to non-noisy strategies that are more likely to cooperate). This implies that any sufficiently small group of mutants who cooperate with positive probability after they observe two or more defections is outperformed.

Next, consider mutants with a small mass $\epsilon << 1$ who defect with a probability of $\alpha > 0$ after they observe $\vec{c}$ (which is the message observed most often when an agent is being matched with a non-noisy incumbent). In what follows we calculate their expected payoff as a function of α in any nearby focal post-entry configuration, neglecting terms of $O\left(\bar{\delta}\right)$ throughout the calculation.[16] Observe that the mutant's partner observes a single defection with a probability of $k \cdot \alpha \cdot (1-\alpha)^{k-1}$, and observes at least two defections with a probability of $1-(1-\alpha)^k - k \cdot \alpha \cdot (1-\alpha)^{k-1}$. This implies that the mean probability that the partner defects against the mutant is:

$$h(\alpha) := \left(k \cdot \alpha \cdot (1-\alpha)^{k-1}\right) \cdot q + 1 - (1-\alpha)^k - k \cdot \alpha \cdot (1-\alpha)^{k-1} = 1 - (1-\alpha)^{k-1}(1 - \alpha + k \cdot \alpha \cdot (1-q)).$$

Thus the expected payoff of the mutant is:

$$\begin{aligned} \pi(\alpha): &= (1-h(\alpha)) \cdot \alpha \cdot (1+g) + (1-h(\alpha)) \cdot (1-\alpha) - h(\alpha) \cdot (1-\alpha) \cdot l \\ &= 1 + \alpha \cdot g - h(\alpha) \cdot (1 + (1-\alpha) \cdot l + \alpha \cdot g). \end{aligned}$$

Direct numeric calculation of $\pi'(\alpha)$ yields that $\pi(\alpha)$ is strictly decreasing in α for each $q > \frac{g}{k\cdot(l+1)}$. Thus the "mutant" with $\alpha = 0$ earns the most, but this is precisely the α of the incumbents.

Next, consider a sufficiently small group of $\epsilon << 1$ mutants who defect (on average) with a probability of $q' \neq q_n$ after observing a single defection (and play the same as the incumbents otherwise). These mutants are strictly outperformed due to the following argument. Recall that q is defined such that both actions are best replies after a player observes a single defection because it balances the direct gain from defecting (which is independent of q_n) and the indirect loss from defecting (which is increasing in q_n). If $q' > q$ ($q' < q_n$), then the average probability in a post-entry focal configuration that a partner defects when the player observes a single defection is $\epsilon \cdot q'_n + (1-\epsilon) \cdot q_n$, which is larger (smaller) than q_n. This implies that the indirect loss of defecting is larger (smaller) than the direct gain, and as a result the mutants who defect with a higher (lower) probability are outperformed.

[16] Nearby focal configurations exist due to the same argument as in the analysis of the noisy strategies above (which show that all the non-noisy strategies defect with a probability of $O(\delta_n)$ as long as $q < 1/k$).

Let χ be the probability that a random partner defects conditional on both the agent and the partner observing a single defection (in the limit as $\delta_n \to 0$):

$$\chi = \lim_{n\to\infty}\left(\sum_{s\in\mathcal{S}} Pr\left(s|d,\overrightarrow{c}\right)\cdot s_{\left(d,\overrightarrow{c}\right)}(d) + Pr\left(\text{TFT}_{q,\xi}|d,\overrightarrow{c}\right)\cdot q\right).$$

We conclude by showing that if $\chi > \mu$ ($\chi < \mu$), then (σ_n, η_n) $((\sigma'_n, \eta'_n))$ is evolutionarily stable. This is so because if $\chi > \mu$ ($\chi < \mu$), then conditional on a non-noisy agent observing a single defection, the partner is more (less) likely to defect the higher the probability with which the agent defects when he observes a single defection (because then it is more likely that the partner observes a single defection rather than only cooperation). This implies that when a player observes a single defection, defection is more (less) profitable the higher the agent's own defection probability is (recall that the direct gain of defection is higher the larger the defection probability of the partner, while the indirect loss is independent of the partner's behavior). That is, an agent's payoff is a strictly convex (concave) function of the agent's defection probability conditional on him observing a single defection. This implies that mutants who mix on the individual level (defect with probabilities different from q) are outperformed when $\chi > \mu$ ($\chi < \mu$)). When $\chi = \mu$, one can show that there is either a sequence of δ_n in which (σ_n, δ_n) is evolutionarily stable or a sequence in which (σ'_n, δ'_n) is evolutionarily stable. □

A.4 Proof of Theorem 4 (Observing Conflicts)

Proof. We first deal with Part 1, namely, the case of a mild PD *($g < \frac{l+1}{2}$).* Recall that under the observation of conflicts, signal D denotes a conflict (at least one player defected), and C denotes mutual cooperation. Let TFT, $TF2T$, TFT_q (and similarly, TFT_ξ, $TF2T_\xi$, $TFT_{q,\xi}$) be defined in an analogous way to the proof of Theorem 3. Let σ be the strategy that assigns mass q (defined below) to TFT and mass $1-q$ to $TF2T$, and let $\sigma' \equiv TFT_q$. Let $\zeta = (\xi, \mathcal{S}, \lambda)$ be a noise structure with a grain of full-support strategy mistakes. Let $\eta = \eta' \equiv c$. Let $\bar{\delta}$ be a number that is sufficiently small relative to $\frac{1}{k\cdot(l+g+1)}$. For each $n \geq 1$, let $\delta_n = \bar{\delta}/n$. Let σ_n (σ'_n) be the closest strategy to σ (σ') in Σ_{ζ,δ_n}.

We now show that there exists sequences of consistent outcomes η_n and η'_n such that η_n, $\eta'_n \to c$ as $n \to \infty$. For each $s \in C(\sigma_n)$ or $s \in C(\sigma'_n)$, let $Pr\left(D, \overrightarrow{C}|s\right)$ and $Pr(D, D, ...|s)$ denote, respectively, the probability of observing exactly one D, and the probability of observing at least two Ds, conditional on the partner following strategy s. Let $Pr\left(D, \overrightarrow{C}\right)$ and $Pr(D, D, ...)$ be the corresponding unconditional probabilities in configuration (σ_n, η_n). Calculations analogous to those explicitly detailed in the proof of Theorem 3 enable us to find $Pr\left(D, \overrightarrow{C}|s\right)$ and $Pr(D, D, ...|s)$ for each strategy s. In particular, as in the previous analysis, the incumbents will very rarely be observed to have two or more conflicts: $Pr(D, D, ...|TFT_{q,\xi}) = O\left(\delta_n^2\right)$. (As in the proof of Theorem 3 we simplify the exposition by letting $TFT_{q,\xi}$ denote the strategy that puts probability q on TFT_ξ and probability $1-q$ on $TF2T_\xi$ in configuration (σ_n, η_n).) Next we calculate the average probability of a player observing a single conflict conditional on the partner following a non-noisy strategy:

$$Pr\left(D, \overrightarrow{C}|\text{TFT}_{q,\xi}\right) = O(\delta_n) + q\cdot 2\cdot k\cdot Pr\left(D, \overrightarrow{C}|\,TFT_{q,\xi}\right) \Rightarrow Pr\left(D, \overrightarrow{C}|\text{TFT}_{q,\xi}\right) = \frac{O(\delta_n)}{1-2\cdot k\cdot q}. \quad (3)$$

This expression is derived as follows. The probability that a non-noisy agent has a conflict with a noisy agent is $O(\delta_n)$ since this is the fraction of noisy agents. A non-noisy agent defects against another non-noisy agent

with an an average probability of q if he observes $D, \overrightarrow{C}$. This observation happens with a probability of $2 \cdot k \cdot Pr\left(D, \overrightarrow{C} | \text{TFT}_{q,\xi}\right) + O\left(\delta_n\right)$ (because in each of the observed k interactions, the two interacting agents each induce a conflict by defecting with an average probability of $Pr\left(D, \overrightarrow{C} | \text{TFT}_{q,\xi}\right)$.

The terms represented by $O\left(\delta_n\right)$ in Eq. 3 are positive. Thus since $Pr\left(D, \overrightarrow{C} | \text{TFT}_{q,\xi}\right) > 0$ it must be the case that

$$2 \cdot k \cdot q < 1 \Leftrightarrow k \cdot q < 0.5.$$

By Bayes' rule we can calculate the conditional probability $Pr\left(s | D, \overrightarrow{C}\right)$ of being matched with each strategy $s \in C\left(\sigma_n\right)$ as a function of the observed message (similar to the calculations detailed in the proof of Theorem 3). Let μ be the probability that a partner defects conditional on the player observing a message $D, \overrightarrow{C}$ about the partner, and conditional on the partner observing the message $\overrightarrow{C}$:

$$\mu = \sum_{s \in \mathcal{S}} Pr\left(s | D, \overrightarrow{C}\right) \cdot s_{\overrightarrow{C}}\left(D\right) + O\left(\delta_n\right). \tag{4}$$

Note that μ is decreasing in q (as a larger q implies a higher probability of $Pr\left(D, \overrightarrow{C} | TFT\right)$). Moreover, as $q \nearrow \frac{1}{2 \cdot k}$ we have $\mu\left(q\right) \searrow 0$ because $Pr\left(D, \overrightarrow{C} | TFT\right)$ "explodes" as we approach the threshold of $k \cdot q = 0.5$.

Next, we calculate the value of q that balances the payoffs of both actions when a player observes a single conflict (neglecting terms of $O\left(\delta_n\right)$). The LHS of the following equation represents a player's direct gain from defecting in the rare case in which he observes a single conflict, while the RHS represents the player's indirect loss from defecting in this case, which is induced by other partners who defect as a result of observing these defections. Note that the cost is paid only if the partner cooperated, as otherwise other partners observe D regardless of the agent's own action.

$$\mu \cdot l + (1 - \mu) \cdot g = (1 - \mu) \cdot k \cdot q \cdot (l + 1) \quad \Leftrightarrow \quad q = \frac{\mu \cdot l + (1 - \mu) \cdot g}{(1 - \mu) \cdot k \cdot (l + 1)}. \tag{5}$$

In connection with Eq. 5 it was noted that $q\left(\mu\right)$ is increasing in μ, and since the PD is mild we have $q\left(0\right) = \frac{g}{k \cdot (l+1)} < \frac{1}{2 \cdot k}$. This implies that there are (unique) values of $\frac{g}{k \cdot (l+1)} < q < \frac{1}{2 \cdot k}$ and $0 < \mu < 1$ that jointly solve Eqs. 4 and 5. By standard continuity arguments, for any n, there exists a nearby frequency $q_n = q + O\left(\delta_n\right)$ that balances the payoffs of the two actions.

The remaining arguments are analogous to those in the final part of the proof of Theorem 3, and are omitted for brevity.

Next, we deal with Part 2, namely, the case of an acute PD *($g > 0.5 \cdot (l + 1)$)*. Cooperation can be perfectly neutrally stable only if non-noisy agents: (1) cooperate with probability one after they observe $\overrightarrow{C}$ (otherwise the outcome cannot converge to full cooperation as the noise converges to zero), and (2) defect (on average) with positive probability after they observe $\left(D, \overrightarrow{C}\right)$. This is because otherwise a mutant who defects with a probability of $0 < \epsilon << 1$, regardless of the observed signal, would earn a direct gain of $O\left(\epsilon\right)$ from defecting, but suffer an indirect loss of at most $O\left(\epsilon^2\right)$ due to these defections (since non-noisy incumbents defect only when they observe at least two conflicts, which happens with a probability of $k \cdot O\left(\epsilon^2\right)$).

Let $q > 0$ denote the average probability of a player defecting after he observes $\left(D, \overrightarrow{C}\right)$. The fact that q is positive implies that defection must be a best reply after a player observes a single conflict. This implies that q should be at least equal to the minimal solution of Eq. (5): $q\left(\mu = 0\right) = \frac{g}{k \cdot (l+1)}$ (assuming that the level of noise is sufficiently low). However, if the game is acute, then this minimal solution is larger than

$\frac{1}{2\cdot k}$. This implies, due to Eq. (3), that an arbitrarily small group of mutants who always defect would cause the incumbents to defect with high probability, which implies that no focal post-entry population exists, and thus cooperation cannot be neutrally stable. □

A.5 Proof of Theorem 5 (Observing Unilateral Defections)

Proof. Let G be a standard PD (i.e., $g < l+1$). Recall that under the observation of unilateral defections, D is the signal for a unilateral defection of the partner, and C is the signal for all other action profiles. Let TFT, $TF2T$, TFT_q (and similarly, TFT_ξ, $TF2T_\xi$, $TFT_{q,\xi}$) be defined in an analogous way to the proof of Theorem 3. Let σ be the strategy that assigns mass q (defined below) to TFT and mass $1-q$ to $TF2T$, and let $\sigma' \equiv TFT_q$. Let $\zeta = (\xi, \mathcal{S}, \lambda)$ be a noise structure with a grain of full-support strategy mistakes. Let $\eta = \eta' \equiv c$. Let $\bar{\delta}$ be a number that is sufficiently small relative to $\frac{1}{k\cdot(l+g+1)}$. For each $n \geq 1$, let $\delta_n = \bar{\delta}/n$. Let σ_n (σ'_n) be the closest strategy to σ (σ') in Σ_{ζ,δ_n}.

We now show the existence of a consistent outcome η_n (η'_n) in which $\eta_n, \eta'_n \to c$ as $n \to \infty$. For each $s \in C(\sigma_n)$ ($s \in C(\sigma'_n)$), let $Pr\left(D, \overrightarrow{C}|s\right)$ and $Pr(D, D, ...|s)$ denote, respectively, the probability of observing exactly one D, and the probability of observing at least two Ds, conditional on the partner following strategy s. Let $Pr\left(D, \overrightarrow{C}\right)$ and $Pr(D, D, ...))$ be the corresponding unconditional probabilities in configuration (σ_n, η_n). Calculations analogous to those explicitly detailed in the proof of Theorem 3 enable us to find $Pr\left(D, \overrightarrow{C}|s\right)$ and $Pr(D, D, ...|s)$ for each strategy s. In particular, as in the previous analysis, the incumbents will very rarely be observed to have two or more Ds: $Pr(D, D, ...|TFT_{q,\xi}) = O\left(\delta_n^2\right)$. Next, we calculate the order of magnitude of $Pr\left(D, \overrightarrow{C}|\, TFT_{q,\xi}\right)$ (the average probability that a player observes a single unilateral defection conditional on the partner following a non-noisy strategy):

$$Pr\left(D, \overrightarrow{C}|\mathrm{TFT}_{q,\xi}\right) = O(\delta_n) + q\cdot k\cdot Pr\left(D, \overrightarrow{C}|\, TFT_{q,\xi}\right) \Rightarrow Pr\left(D, \overrightarrow{C}|\mathrm{TFT}_{q,\xi}\right) = \frac{O(\delta_n)}{1-k\cdot q}. \tag{6}$$

This expression is derived as follows. The probability that a non-noisy agent unilaterally defects against a noisy agent is $O(\delta_n)$ since this is the fraction of noisy agents. A non-noisy agent defects against another non-noisy agent with an average probability of q if he observes $\left(D, \overrightarrow{C}\right)$ (and this defection is a unilateral defection with a probability of $1-O(\delta_n)$). This observation happens with a probability of $k\cdot Pr\left(D, \overrightarrow{C}|\mathrm{TFT}_{q,\xi}\right)+O(\delta_n)$ (because in each of the observed k interactions, the partner unilaterally defects with an average probability of $Pr\left(D, \overrightarrow{C}|\mathrm{TFT}_{q,\xi}\right)$.

The terms represented by $O(\delta_n)$ in Eq. 6 are positive. Thus since $Pr\left(D, \overrightarrow{C}|\mathrm{TFT}_{q,\xi}\right) > 0$ we have

$$Pr\left(D, \overrightarrow{C}|\, TFT_{q,\xi}\right) = O(\delta_n) \Leftrightarrow k\cdot q < 1 \Leftrightarrow k\cdot q < 1.$$

By using Bayes' rule we can calculate the conditional probability $Pr\left(s|D, \overrightarrow{C}\right)$ for being matched with each strategy $s \in C(\sigma_n)$ as a function of the observed message (explicit calculations were presented in the proof of Theorem 3). Let μ be the probability that a partner defects conditional on the player observing message $\left(D, \overrightarrow{C}\right)$ about the partner, and conditional on the partner observing the message $\overrightarrow{C}$:

$$\mu = \sum_{s\in\mathcal{S}} Pr\left(s|D, \overrightarrow{C}\right)\cdot s_{\overrightarrow{C}}(D) + O(\delta_n). \tag{7}$$

Note that μ is decreasing in q (as a larger q implies a larger $Pr\left(D,\overrightarrow{C}\,|\,TFT_{q,\xi}\right)$). Moreover, as $q \nearrow \frac{1}{k}$, $\mu(q) \searrow 0$ because $Pr\left(D,\overrightarrow{C}\,|\,TFT_{q,\xi}\right)$ "explodes" as $k\cdot q \nearrow 1$.

Next, we calculate the value of q that balances the payoffs of both actions when a player observes a single unilateral defection (neglecting terms of $O(\delta_n)$). The LHS of the Eq. (8) presents the direct gain from defecting in these cases, while the RHS presents the indirect loss from these defections as a result of inducing other partners who observe these interactions to defect. Observe that the cost is paid only if the partner cooperates, as otherwise the signal C would be observed regardless of the agent's action.

$$\mu\cdot l+(1-\mu)\cdot g=(1-\mu)\cdot k\cdot q\cdot(l+1)\;\;\Leftrightarrow\;\;q=\frac{\mu\cdot l+(1-\mu)\cdot g}{(1-\mu)\cdot k\cdot(l+1)}. \tag{8}$$

Observe, that $q(\mu)$ is increasing in μ, and $q(0)=\frac{g}{k\cdot(l+1)}<\frac{1}{k}$ (due to the PD being "standard"). This implies that there are (unique) values of $\frac{g}{k\cdot(l+1)}<q<\frac{1}{k}$ and $0<\mu<1$ that jointly solve Eqs. 7 and 8. By standard continuity arguments, for any n, there exists a nearby frequency $q+O(\delta_n)$ that balances the payoff of the two strategies. The remaining arguments are analogous to those in the last part of the proof of Theorem 3, and are omitted for brevity.

Next, we deal with Part 2, namely, the case of an inefficient PD *($g>l+1$)*. Assume to the contrary, that cooperation is a perfectly neutrally stable outcome. Cooperation can be the outcome of the limit of the perfectly neutrally stable configurations only if non-noisy agents cooperate with probability one after they observe $\overrightarrow{C}$. Moreover, the stability of cooperation requires that the non-noisy agents defect (on average) with positive probability after they observe $\left(D,\overrightarrow{C}\right)$ (otherwise mutants who defect with a probability of $0<\epsilon<<1$, regardless of the observed signal, would earn a direct gain of $O(\epsilon)$ from defecting, but suffer a smaller indirect loss of at most $k\cdot O\left(\epsilon^2\right)$ due to these defections).

Let q denote the average probability of defection after players observe $\left(D,\overrightarrow{C}\right)$. The fact that non-noisy agents defect with positive probability after observing $\left(D,\overrightarrow{C}\right)$ implies that cooperation should be a best reply when a player who almost always cooperates observes $\left(D,\overrightarrow{C}\right)$. This implies that q should be at least equal to the minimal solution of Eq. (5): $q(\mu=0)=\frac{g}{k\cdot(l+1)}$ (assuming that the level of noise is sufficiently low). However, if the game is inefficient, then the minimal solution of the equation, $\frac{g}{k\cdot(l+1)}>\frac{1}{k}$, which implies by Eq. (3) that an arbitrarily small group of mutants who defect with small probability would cause the incumbents to unilaterally defect with high probability, and thus no focal post-entry population would exist, which contradicts the assumption that cooperation is perfectly neutrally stable. □

A.6 Proof of Theorem 6 (Observing Action Profiles)

Proof. We begin with case 1, in which G is a mild PD. Let $\zeta=(\xi,\mathcal{S},\lambda)$ be a noise structure in which $\xi\equiv 0$,[17] where $\mathcal{S}$ contains a single strategy s_α that defects with a small probability of $0<\alpha<<\frac{1}{k}$ regardless of the observed signal.[18] Let TF2T be the strategy that defects iff the observed message includes at least two interactions in which the action profile is different from mutual cooperation. Let TFT be the strategy that defects if the observed message includes either (1) at least two interactions in which the action profile is different from mutual cooperation, or (2) at least one interaction in which the partner was the sole defector.

[17] The assumption that $\xi\equiv 0$ is taken to simplify the arguments, but it does not play an essential role in the proof.

[18] In order to satisfy the requirement of Definition 4 that $\mathcal{S}$ includes two different strategies with different defection probability, we can slightly adapt the construction and have $\mathcal{S}$ to include two noisy strategies, such that the first (second) strategy defects with a probability of $1.001\cdot\alpha$ ($0.999\cdot\alpha$).

Let σ^* be the strategy that assigns mass q (defined below) to TFT and mass $1-q$ to TF2T. Let $\eta^* \equiv c$. Let $\bar{\delta}$ be a number that is sufficiently small relative to $1/(k\cdot(l+g+1))$. For each $n \geq 1$, let $\delta_n = \bar{\delta}/n$. Let $\sigma_n \in \Sigma_{\zeta,\delta_n}$ be the closest strategy to σ^* in Σ_{ζ,δ_n}. We now show the existence of a consistent outcome η_n in which $\eta_n \to \eta^* \equiv c$ as $n \to \infty$.

Let $Pr\left(s_\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)$ be the probability that the partner follows strategy s_α conditional on the player observing a signal profile with a single unilateral defection by the partner, and $k-1$ mutual cooperations. Let μ be the probability that the partner of a non-noisy agent defects conditional on the agent observing $\left(\left(d,c\right),\overrightarrow{(c,c)}\right)$. Note, that $\mu = \alpha\cdot Pr\left(s_\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)+O\left(\delta_n\right)$ because the non-noisy strategies cooperate with maximal probability upon observing $\left(\overrightarrow{(c,c)}\right)$. The value of q is defined to make an agent, who almost always cooperates, indifferent between cooperation and defection when he observes message $\left(\left(d,c\right),\overrightarrow{(c,c)}\right)$, namely:

$$\mu\cdot l+(1-\mu)\cdot g=(1-\mu)\cdot k\cdot q\cdot(l+1) \;\Leftrightarrow\; q=\frac{\mu\cdot l+(1-\mu)\cdot g}{(1-\mu)\cdot k\cdot(l+1)}. \tag{9}$$

For a sufficiently small α, the value of $q\left(\mu\right)$ that solves Eq. (9) will be slightly above $q\left(\mu=0\right)=\frac{g}{l+1}$. The fact that the PD is mild implies that (for a sufficiently small α) $k\cdot q<0.5$.

Let p be the average probability with which the non-noisy players defect when being matched with s_α. When $\alpha << \frac{1}{k}$, the TF2T agents rarely $(O\left(\alpha^2\right))$ defect against the noisy agents, because it is rare to observe them defecting more than once. The TFT agents defect against the s_α-agents with a probability of $k\cdot q\cdot\alpha+O\left(\alpha^2\right)$ because each rare defection of the s-agents is observed with a probability of $k\cdot q$ by TFT-agents. As both $\alpha, p << 1$, it implies that bilateral defections are very rare $((O\left(\alpha^2\right))$. This implies that $p=\alpha\cdot k\cdot q+O\left(\alpha^2\right)<\frac{\alpha}{2}$.

Let r be the probability that a TFT agent defects against a fellow TFT agent. In each observed interaction, the TFT partner interacts with a noisy (resp., TFT, TF2T) opponent with a probability of δ_n (resp., q, 1-q) and the partner unilaterally defects with a probability of $\alpha\cdot k\cdot q+O\left(\alpha^2\right)$ (resp., $r+O\left(r^2\right)$, $O\left(\delta_n\cdot\alpha^2\right)$). This implies that r solves the following equation:

$$r=k\cdot(\alpha\cdot q\cdot\delta_n+q\cdot r)+O\left(\delta_n^2\right) \;\Rightarrow\; r=\frac{\alpha\cdot k\cdot q}{1-k\cdot q}\cdot\delta_n+O\left(\delta_n^2+\alpha^2\cdot\delta_n\right)<0.5\cdot\alpha\cdot\delta_n,$$

where the latter inequality is because $k\cdot q<0.5$. The above calculations show that the total frequency with which noisy agents unilaterally defect $(\alpha\cdot\delta_n)$ is higher than the total frequency with which non-noisy agents defect $(q+p\cdot\delta_n<\alpha\cdot\delta_n)$. This implies that the probability that an agent is noisy, conditional on his being the sole defector in an interaction, is higher than 50%, and that it is larger than this probability conditional on his being the sole cooperator. Next, note that mutual defections between a noisy and a TFT agent have a frequency of $O\left(\delta_n\right)$, while mutual defections between two noisy agents (or two non-noisy agents) are very rare $(O\left(\delta_n^2\right))$, which implies that the probability that the partner follows a noisy strategy conditional on the player observing mutual defection is 50%+$O\left(\delta_n\right)$. This implies that

$$Pr\left(s_\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)>\max\left(Pr\left(s_\alpha|\left(d,d\right),\overrightarrow{(c,c)}\right)>Pr\left(s_\alpha|\left(c,d\right),\overrightarrow{(c,c)}\right)\right),$$

and thus while both actions are best replies after the player observes the message $\left(\left(d,c\right),\overrightarrow{(c,c)}\right)$, only cooperation is a best reply after the player observes $\left(\left(d,d\right),\overrightarrow{(c,c)}\right)$ and $\left(\left(c,d\right),\overrightarrow{(c,c)}\right)$. Next note that conditional on a player observing a message with at most $k-2$ mutual cooperations, the partner is most likely to be

a noisy agent (because non-noisy agents have two outcomes different from mutual cooperation with a probability of $O\left(\delta_n^2\right)$. This implies that the non-noisy agents play the unique best-reply after any signal other than $\left((d,c),\overrightarrow{(c,c)}\right)$, and thus any small group of mutants who behave differently in these cases will be outperformed. The stability with respect to mutants who differ in their behavior after $\left((d,c),\overrightarrow{(c,c)}\right)$ is derived by analogous arguments as in the end of the proof of Theorem 3.

Next we show that cooperation is not perfectly neutrally stable for all (some) noise structures in acute (mild) PDs. Note that both the direct payoff and the indirect payoff of an action (the latter payoff being due to the influence of the action on the behavior of other partners) depend only on the conditional probability that the partner defects. In order to support stable cooperation, cooperation (defection) should be the unique best reply against a partner who is going to cooperate (defect) for sure, and both actions should be best replies to some conditional probability strictly between zero and one. Moreover, non-noisy agents should defect with positive probability when they observe $\left((d,c),\overrightarrow{(c,c)}\right)$ (as otherwise defecting with a probability of $0<\epsilon<<1$ against cooperative partners would be profitable). This can be the case only if conditional on a player observing $\left((d,c),\overrightarrow{(c,c)}\right)$ there is a positive probability that the partner follows a noisy strategy (and that this probability is higher than the conditional probability when the player observes $\left(\overrightarrow{(c,c)}\right)$. Note also that all non-noisy agents must defect with probability one when they observe at least two interactions with outcomes different from mutual cooperation, (because then it is most likely that the partner is a noisy agent, and the conditional probability that the partner defects is higher than when the player observes $\left((d,c),\overrightarrow{(c,c)}\right)$).

We first show that there exists a noise structure ζ such that cooperation is not perfectly neutrally stable with respect to ζ for any mild PD. Let $\zeta=(\xi,\mathcal{S},\lambda)$ be a noise structure in which: (1) $\xi\equiv 0$ (but the proof can be adapted to $\xi>0$), and (2) all the noisy strategies in $\mathcal{S}$ defect with a probability higher than $\frac{2}{3}$ regardless of the observed signal. In what follows, we show that the non-noisy players defect with probability one against those non-noisy agents. Let $s_\alpha\in S$ be a noisy strategy that defects with a probability of α regardless of the signal. The following inequality bounds $1-p$ from above:

$$1-p \quad\leq\quad \left((1-\alpha)\cdot(1-p)\right)^k+k\cdot\left((1-\alpha)\cdot(1-k)\right)^{k-1}\cdot\left(1-(1-\alpha)\cdot(1-p)\right),$$

because a non-noisy agent cooperates (the LHS) only if observes at least $k-1$ mutual cooperations (the RHS). If $p<1$ we can divide by $1-p$ and get:

$$1\leq(1-\alpha)^k\cdot(1-p)^{k-1}+k\cdot(1-\alpha)^{k-1}\cdot(1-k)^{k-1}\cdot\left(1-(1-\alpha)\cdot(1-p)\right).$$

Note that $\alpha>\frac{2}{3}$ implies that $(1-\alpha)^k<\frac{1}{9}$ and $k\cdot(1-\alpha)^{k-1}<\frac{2}{3}$, which implies that the RHS is less than 1, and we get a contradiction. Thus it must be that $p=1$; i.e., the non-noisy players always defect against the noisy agents. This implies that the probability that the partner is a noisy player conditional on the player observing $\left((d,c),\overrightarrow{(c,c)}\right)$ is zero, and we get a contradiction.

Now we deal with case 2, in which the PD is acute, and the noise structure is arbitrary. Assume to the contrary, that cooperation is a perfectly neutrally stable outcome. Let $\zeta=(\xi,\mathcal{S},\lambda)$ be the supporting noise structure (with a grain of full-support strategy mistakes). For each noise level δ_n, let $q_n>0$ be the average probability according to which a non-noisy incumbent defects when he observes $\left((d,c),\overrightarrow{(c,c)}\right)$. Let q be the limit of q_n when δ_n converges to zero. Eq. (2) and the arguments associated with it show that $k\cdot q>\frac{g}{(l+1)}$

is necessary for cooperation to be a best reply for a player who observes $\left(\overrightarrow{(c,c)}\right)$. Recall that in acute PDs $\frac{g}{(l+1)} > \frac{1}{2} \Rightarrow k \cdot q > \frac{1}{2}$.

Let $s_\alpha \in \mathcal{S}$ be a noisy strategy that induces an agent who follows it (called s_α-agent) to defect with a probability of $\alpha > 0$ when he observes $\left(\overrightarrow{(c,c)}\right)$. In what follows we show that the presence of strategy s_α induces the non-noisy agents to unilaterally defect more often than s_α-agents do so. Let p be the average probability that non-noisy agents defect against s_α-agents. This probability p must solve the following inequality:

$$\begin{aligned} 1-p \quad \geq \quad & \left((1-\alpha)\cdot(1-k)\right)^k + k\cdot\left((1-\alpha)\cdot(1-k)\right)^{k-1}\cdot\left(1-(1-\alpha)\cdot(1-p)\right) \\ & +(1-q)\cdot k\cdot\left((1-\alpha)\cdot(1-k)\right)^{k-1}\cdot\alpha\cdot(1-p)\,. \end{aligned} \tag{10}$$

The LHS of (10) is the average probability that non-noisy agents cooperate against s_α-agents (recall that non-noisy agents always defect when they observe less than $k-1$ mutual cooperations). The non-noisy agents cooperate with probability one (resp., at most one, q) if they observe $\left(\overrightarrow{(c,c)}\right)$ (resp., $\left((d,d),\overrightarrow{(c,c)}\right)$ or $\left((c,d),\overrightarrow{(c,c)}\right)$, $\left((d,c),\overrightarrow{(c,c)}\right)$), which happens with a probability of $\left((1-\alpha)\cdot(1-k)\right)^k$ (resp., $k\cdot\left((1-\alpha)\cdot(1-k)\right)^{k-1}\cdot\left(1-(1-\alpha)\cdot(1-p)\right)$, $k\cdot\left((1-\alpha)\cdot(1-k)\right)^{k-1}\cdot\alpha\cdot(1-p)$).

Direct numerical analysis of Eq. (10) shows that the minimal p that solves this inequality (given that $q > \frac{1}{2\cdot k}$) is greater than $\frac{\alpha}{2-\alpha}$ for any $0<\alpha<1$. The total frequency of interactions in which the s_α-agents unilaterally defect is $\alpha\cdot(1-p)\cdot\delta\cdot\lambda(s)$. The total frequency of interactions in which non-noisy agents unilaterally defect against the s_α-agents is $p\cdot(1-\alpha)\cdot\delta\cdot\lambda(s)$. Eq. (7) shows that these unilateral defections against s_α-agents induce the non-noisy agents to unilaterally defect among themselves with a total frequency of $\frac{p\cdot(1-\alpha)\cdot\delta\cdot\lambda(s)}{1-k\cdot q} > p\cdot(1-\alpha)\cdot\delta\cdot\lambda(s)$. Finally, note that $p > \frac{\alpha}{2-\alpha} \Leftrightarrow 2\cdot p\cdot(1-\alpha) > \alpha\cdot(1-p)$ implies that non-noisy agents unilaterally defect (as the indirect result of the presence of the s-agents) more often than those in which the s_α-agents do.

Next, observe that bilateral defections are most likely to occur in interactions between noisy and non-noisy agents. This is because the probability that both non-noisy agents defect against each other is only $O\left(\delta^2\right)$. Thus, when a player observes bilateral defection the partner is more likely to be a noisy agent than when the player observes a unilateral defection by the partner. This implies that all the non-noisy agents defect with probability one when they observe $\left((d,d),\overrightarrow{(c,c)}\right)$ because in this case defection is the unique best reply.

Let w be the (average) probability that non-noisy agents defect when they observe $\left((c,d),\overrightarrow{(c,c)}\right)$. If $w<0.5$, then cooperation is the unique best reply for a non-noisy agent who faces a partner who is likely to defect (e.g., when they observe fewer than $k-1$ mutual cooperations), and so we get a contradiction. This is because defecting against a defector yields a direct gain of l and an indirect loss of at least $0.5\cdot k\cdot(l+1) \geq l+1 > l$ (because this bilateral defection will be observed on average k times, and in at least half of these cases it will induce the partner to defect, whereas if the agent were cooperating, then he would have induced the partner to cooperate).

Thus, $w \geq 0.5 \Rightarrow k\cdot w > 1$. However, in this case, analogous arguments to those after Eq. (6) imply that an arbitrarily small group of mutants who defect with small probability would cause the incumbents to unilaterally defect with high probability, and thus no focal post-entry population exists, which contradicts the assumption that cooperation is neutrally stable. □

A.7 Proof of Prop. 1 (Strict Equilibrium is Perfectly Stable)

Proof. Let $\bar{k} = \text{argmax}\left(C\left(p\right)\right)$. Let l be the minimal loss from playing $a \neq a^*$ against a^*: $l = \min_{a \neq a^*}\left(\pi\left(a^*, a^*\right) - \pi\left(a, a^*\right)\right)$. Let g be the highest possible payoff in the game: $g = \max_{a,a'} \pi\left(a, a'\right)$.

Let τ be sufficiently small with respect to $\frac{l}{g}$. Let $\hat{s}$ be the strategy that plays a^* with a probability of $1-\tau$, and plays each other action with a probability of $\frac{\tau}{|A|}$. Let $\zeta = \left(\xi = 0, \mathcal{S} = \left\{\hat{s}\right\}, \lambda = 1\right)$ be a noise structure that includes a single source of noise: the strategy $\hat{s}$ that plays a^* with high probability. Let $\left(\delta_n\right)_n \to 0$ be any sequence of noise level converging to 0. Let σ_n be the closest strategy to a^* in Σ_{ζ,δ_n}: $\sigma_n\left(a^*\right) = 1 - \delta \cdot \lambda$. Let η_n be the unique consistent outcome of σ_n. It is immediate that $\left(\sigma_n, \eta_n\right) \to \left(d, d\right)$. Fix n. We have to show that $\left(\sigma_n, \eta_n\right)$ is an evolutionarily stable configuration in the perturbed environment $\left(E, p, \zeta, \delta_n\right)$.

Let $\bar{\epsilon}$ be sufficiently small with respect to $\frac{l}{\bar{k} \cdot g}$. Let $0 < \epsilon < \bar{\epsilon}$. Let $\sigma' \neq \sigma_n \in \Sigma_{\zeta,\delta_n}$ be a mutant strategy. Let $\left(\sigma_\epsilon, \eta_\epsilon\right)$ be an ϵ-post-entry configuration. It is immediate that $\left(\sigma_\epsilon, \eta_\epsilon\right)$ is 0-focal because all the incumbents play the same regardless of the observed message. We have to show that the mutants are strictly outperformed. The mutants play a^* with a strictly lower probability than the incumbents (because $\sigma' \neq \sigma_n$ and all messages are observed on the equilibrium path). Let β be the difference between the probability of playing a^* by the mutants and by the incumbents, when facing an incumbent. An analogous argument to the one in Theorem 1 above shows that the maximal probability, $\bar{\alpha}$, that one mutant plays an action different from a^* against another mutant is at most

$$\bar{\alpha} \leq \bar{k} \cdot \left(\beta + \epsilon \cdot 2 \cdot \bar{\alpha}\right) \Rightarrow \bar{\alpha} \leq \frac{\bar{k} \cdot \beta}{1 - \epsilon \cdot 2}.$$

The mutants suffer a direct loss of $\beta \cdot l$ from their lower probability of playing a^* against the incumbents. Their indirect gain (from inducing other mutants to play more favorably towards them) is at most $\epsilon \cdot \bar{\alpha} \cdot g$. Thus the loss outweighs the gain if:

$$\epsilon \cdot \bar{\alpha} \cdot g = \frac{\bar{k} \cdot \beta \cdot \epsilon \cdot g}{1 - \epsilon \cdot 2} < \beta \cdot l,$$

which holds for our choice of $\bar{\epsilon}$ as sufficiently small. □

References

Abreu, D., and R. Sethi (2003): "Evolutionary stability in a reputational model of bargaining," *Games and Economic Behavior*, 44(2), 195–216.

Alger, I., and J. W. Weibull (2013): "Homo Moralis - Preference Evolution Under Incomplete Information and Assortative Matching," *Econometrica*, 81(6), 2269–2302.

Berger, U., and A. Grüne (2014): "Evolutionary Stability of Indirect Reciprocity by Image Scoring," Mimeo.

Bernstein, L. (1992): "Opting out of the legal system: Extralegal contractual relations in the diamond industry," *The Journal of Legal Studies*, pp. 115–157.

Björnerstedt, J., and J. Weibull (1996): "Nash equilibrium and evolution by imitation," in *The Rational Foundations of Economic Behaviour*, ed. by K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt, pp. 155–171. MacMillan, London.

Blonski, M., P. Ockenfels, and G. Spagnolo (2011): "Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence," *American Economic Journal: Microeconomics*, 3(3), 164–192.

Bolton, G. E., E. Katok, and A. Ockenfels (2005): "Cooperation among strangers with limited information about reputation," *Journal of Public Economics*, 89(8), 1457–1468.

Breitmoser, Y. (2015): "Cooperation, but no reciprocity: Individual strategies in the repeated Prisoner's Dilemma," *American Economic Review*, forthcoming.

Dal Bó, P., and G. R. Fréchette (2011): "The evolution of cooperation in infinitely repeated games: Experimental evidence," *The American Economic Review*, 101(1), 411–429.

——— (2015): "Strategy choice in the infinitely repeated prisoners dilemma," SSRN 2292390.

Deb, J. (2012): "Cooperation and Community Responsibility: A Folk Theorem for Repeated Random Matching Games," Mimeo.

Dekel, E., J. C. Ely, and O. Yilankaya (2007): "Evolution of preferences," *The Review of Economic Studies*, 74(3), 685–704.

Dixit, A. (2003): "On modes of economic governance," *Econometrica*, 71(2), 449–481.

Duffy, J., and J. Ochs (2009): "Cooperative behavior and the frequency of social interaction," *Games and Economic Behavior*, 66(2), 785–812.

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

Engelmann, D., and U. Fischbacher (2009): "Indirect reciprocity and strategic reputation building in an experimental helping game," *Games and Economic Behavior*, 67(2), 399–407.

Fujiwara-Greve, T., and M. Okuno-Fujiwara (2009): "Voluntarily separable repeated prisoner's dilemma," *The Review of Economic Studies*, 76(3), 993–1021.

Gong, B., and C.-L. Yang (2010): "Reputation and cooperation: An experiment on prisoner dilemma with second-order information," *Available at SSRN 1549605*.

Greif, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *The American Economic Review*, pp. 525–548.

Güth, W., and M. Yaari (1992): "Explaining reciprocal behavior in simple strategic games: An evolutionary approach," in *Explaining Process and Change: Approaches to Evolutionary Economics*, ed. by U. Witt. University of Michigan Press, Ann Arbor.

Heller, Y. (2015a): "Instability of Equilibria with Private Monitoring," Mimeo.

——— (2015b): "Three steps ahead," *Theoretical Economics*, 10, 203–241.

Heller, Y., and E. Mohlin (2015a): "Coevolution of Deception and Preferences: Darwin and Nash Meet Machiavelli," Mimeo.

——— (2015b): "Unique Stationary Behaivor," Mimeo.

Herold, F. (2012): "Carrot or Stick? The Evolution of Reciprocal Preferences in a Haystack Model," *American Economic Review*, 102(2), 914–40.

Herold, F., and C. Kuzmics (2009): "Evolutionary stability of discrimination under observability," *Games and Economic Behavior*, 67(2), 542–551.

Jøsang, A., R. Ismail, and C. Boyd (2007): "A survey of trust and reputation systems for online service provision," *Decision support systems*, 43(2), 618–644.

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

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27(2), 245–252.

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

Maynard Smith, J., and G. R. Price (1973): "The logic of animal conflict," *Nature*, 246, 15.

Milgrom, P., D. C. North, and B. R. Weingast (1990): "THE ROLE OF INSTITUTIONS IN THE REVIVAL OF TRADE: THE LAW MERCHANT, PRIVATE JUDGES, AND THE CHAMPAGNE FAIRS," *Economics and Politics*, 2(1), 1–23.

Milinski, M., D. Semmann, T. C. Bakker, and H.-J. Krambeck (2001): "Cooperation through indirect reciprocity: image scoring or standing strategy?," *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1484), 2495–2501.

Molander, P. (1985): "The optimal level of generosity in a selfish, uncertain environment," *Journal of Conflict Resolution*, 29(4), 611–618.

Nowak, M. A., and K. Sigmund (1992): "Tit for tat in heterogeneous populations," *Nature*, 355(6357), 250–253.

——— (1998): "Evolution of indirect reciprocity by image scoring," *Nature*, 393(6685), 573–577.

——— (2005): "Evolution of indirect reciprocity," *Nature*, 437(7063), 1291–1298.

Ohtsuki, H., and Y. Iwasa (2006): "The leading eight: social norms that can maintain cooperation by indirect reciprocity," *Journal of Theoretical Biology*, 239(4), 435–444.

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

Okuno-Fujiwara, M., and A. Postlewaite (1995): "Social norms and random matching games," *Games and Economic Behavior*, 9(1), 79–109.

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

Resnick, P., and R. Zeckhauser (2002): "Trust among strangers in internet transactions: Empirical analysis of ebay reputation system," *The Economics of the Internet and E-commerce*, 11(2), 23–25.

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

Robson, A. J., and L. Samuelson (2010): "The evolutionary foundations of preferences," *Handbook of Social Economics, Amsterdam: North Holland*.

Rosenthal, R. W. (1979): "Sequences of games with varying opponents," *Econometrica*, pp. 1353–1366.

Samuelson, L. (1998): *Evolutionary games and equilibrium selection*, vol. 1. Mit Press.

Seinen, I., and A. Schram (2006): "Social status and group norms: Indirect reciprocity in a repeated helping experiment," *European Economic Review*, 50(3), 581–602.

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

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

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

Stahl, D. O. (2013): "An experimental test of the efficacy of a simple reputation mechanism to solve social dilemmas," *Journal of Economic Behavior & Organization*, 94, 116–124.

Sugden, R. (1986): *The Economics of Rights, Co-operation and Welfare*. Blackwell Oxford.

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

van Veelen, M., J. García, D. G. Rand, and M. A. Nowak (2012): "Direct reciprocity in structured populations," *Proc. of the National Academy of Sciences*, 109(25), 9929–34.

Wedekind, C., and M. Milinski (2000): "Cooperation through image scoring in humans," *Science*, 288(5467), 850–852.

Wiseman, T., and O. Yilankaya (2001): "Cooperation, secret handshakes, and imitation in the prisoners' dilemma," *Games and Economic Behavior*, 37(1), 216–242.