Observations on Cooperation

Heller, Yuval and Mohlin, Erik

University of Oxford

19 July 2015

Observations on Cooperation

Yuval Heller* and Erik Mohlin‡

April 12, 2016

Abstract

We study environments in which agents are randomly matched to play a Prisoner's Dilemma, and each player observes a few of the partner's past actions against previous opponents. We depart from the existing literature in two key respects: (1) we allow a small fraction of the population to be commitment types, and (2) we do not assume a time zero at which the entire community starts to interact. The presence of committed agents destabilizes all previously proposed mechanisms for sustaining cooperation. We present a novel mechanism (involving an essentially unique strategy combination) that sustains stable cooperation in many environments.

JEL Classification: C72, C73, D83. **Keywords**: Community enforcement; indirect reciprocity; random matching; Prisoner's Dilemma; image scoring.

1 Introduction

Consider the following example of a simple yet fundamental economic interaction. Alice has to trade with another agent, Bob, whom she does not know. Both sides have opportunities to cheat, to their own benefit, at the expense of the other. Alice is unlikely to interact with Bob again, and thus her ability to retaliate, in case Bob acts opportunistically, is restricted. The effectiveness of external enforcement is also limited, e.g., due to incompleteness of contracts, non-verifiability of information, and court costs. Thus cooperation may be impossible to achieve. Alice asks a couple of her friends, who happen to have interacted with Bob in the past, about Bob's behavior, and she considers this information when she decides how to act. Alice also takes into account that her behavior towards Bob in the current interaction may be observed by her future partners. Historically, the described situation was a challenge to the establishment of long-distance trade (Milgrom, North, and Weingast, 1990; Greif, 1993), and it continues to play an important role in the modern economy, in both offline (Bernstein, 1992; Dixit, 2003) and online interactions (Resnick and Zeckhauser, 2002; Jøsang, Ismail, and Boyd, 2007).

Several papers have studied the question of how cooperation can be supported by means of community enforcement. Our modeling approach differs from the existing literature in two key respects. First, we allow

*Affiliation: Department of Economics and Queen's College, University of Oxford, UK. E-mail: yuval.heller@economics.ox.ac.uk.
†Affiliation: Department of Economics, Lund University, Sweden. E-mail: erik.mohlin@nek.lu.se.

‡A previous version of this paper was circulated under the title "Stable observable behavior." We have benefited greatly from discussions with Vince Crawford, Eddie Dekel, Christoph Kuzmics, Ariel Rubinstein, Larry Samuelson, Bill Sandholm, Rann Smorodinsky, Rani Spiegler, Balázs Szentes, Satoru Takahashi, Jörgen Weibull, and Peyton Young. We would like to express our deep gratitude to seminar/workshop participants at the University of Amsterdam (CREED), University of Bamberg, Bar Ilan University, Bielefeld University, University of Cambridge, Hebrew University of Jerusalem, Helsinki Center for Economic Research, Interdisciplinary Center Herzliya, Israel Institute of Technology, Lund University, University of Oxford, University of Pittsburgh, Stockholm School of Economics, Tel Aviv University, NBER Theory Workshop at Wisconsin-Madison, KAEA session at the ASSA 2015, and the Biological Basis of Preference conference at Simon Fraser University, for many useful comments. Last but not least, we thank Renana Heller for suggesting the title.

a few agents in the population to be committed to behaviors that do not necessarily maximize their payoffs. It turns out that this small perturbation completely destabilizes existing mechanisms to sustain cooperation. Specifically, both the contagious equilibria (Kandori, 1992; Ellison, 1994)[1], and the "belief-free" equilibria (Takahashi, 2010; Deb, 2012)[2] fail in the presence of a small fraction of committed agents.

Our second main departure from the literature is to relax the standard assumption that there is an initial time zero at which the entire community starts to interact (see, e.g., Kandori, 1992; Ellison, 1994; Dixit, 2003; Takahashi, 2010; Deb, 2012; Deb and González-Díaz, 2014). In many real-life situations, the interactions within a community have been going on from time immemorial. Consequently the participants may have only a vague idea of the starting point. It seems implausible that agents would be able to condition their behavior on everything that has happened since then (or on "calendar time"). A major methodological contribution of this paper is therefore the development of a novel methodology for how to analyze steady states of community interactions that do not have a global starting time (and, therefore, are not repeated games). A detailed discussion of this issue and its relation to the existing literature appears in Section 2.6.

Our key results are as follows. First, we show that always defecting is the unique perfect equilibrium, regardless of the number of observed actions, provided that the bonus of defection in the underlying Prisoner's Dilemma is larger when the partner cooperates than when the partner defects. This anti-folk theorem result is striking since several papers have presented folk theorems or claimed that evolution can lead to cooperation in related setups. Second, in the opposite case, when the bonus of defection is larger when the partner defects than when the partner cooperates, we present a novel and essentially unique combination of strategies that sustains cooperation: all agents cooperate when they observe no defections and defect when they observe at least two defections. Some of the agents also defect when observing a single defection. Importantly, this cooperative behavior is robust to many kinds of perturbations, and it appears consistent with experimental data. Third, we extend the model to environments in which an agent also obtains information about the behavior of past opponents against the current partner. We show that in this setup cooperation can be sustained if and only if the bonus of defection of a player is less than half the loss she induces to a cooperative partner. Finally, we characterize an observation structure that allows cooperation to be supported as a perfect equilibrium action in *all* Prisoner's Dilemma games. In all observation structures we use the same essentially unique construction to sustain cooperation. It relies only on simple and empirically plausible strategies. Importantly, it does not require the existence of a third party that provides information about reputations.

Overview of the Model Agents in an infinite population are randomly matched into pairs to play a symmetric one-shot game. Before playing the game, each agent privately draws a random sample of k actions that have been played by her partner against other opponents in the past.[3] Specifically, when the underlying game is the Prisoner's Dilemma, we interpret the observed signal, m, as the number of times the partner defected in the sampled k interactions.

We require each agent to follow a *stationary strategy*: a mapping that assigns a mixed action to each signal

[1] In contagious equilibria players start by cooperating. If one player defects at stage t, her partner defects at period $t+1$, infecting another player who defects at period $t+2$, and so on. The non-robustness of these equilibria to a single "crazy" agent was already noted by Ellison (1994, p. 578): "If one player were 'crazy' and always played D (or simply was unaware which equilibrium was being played) again the contagious strategies would not support cooperation. In large populations, the assumption that all players are rational and know their opponents' strategies may be both very important to the conclusions and fairly implausible."

[2] In belief-free equilibria players are always indifferent between their actions, but they choose different mixed actions depending on the signal they obtain about the partner. To the best of our knowledge we are the first to show the non-robustness of these equilibria to the presence of a few committed agents. Elsewhere, one of us develops a somewhat related critique on belief-free equilibria in a standard setup of repeated games between the same two players (Heller, 2015a).

[3] In the main model these k actions are sampled from the entire history of play of the partner. In Appendix A, we present a variant of the model in which each agent observes the most recent k actions of the partner.

that the agent may observe about the current partner. (That is, the action is not allowed to depend on calendar time or on the agent's own history.)[4] A *steady state* of the environment is a triple: (1) a finite set of strategies S^* that are played by the agents in the population, (2) a distribution σ^* over S^* that describes the fractions of the population following the different strategies, and (3) a behavior mapping η^* that describes the mixed action played by each agent (Alice) who follows strategy $s \in S^*$ conditional on being matched against an agent (Bob) who follows strategy $s' \in S^*$. The mapping η is required to be *consistent* in the sense that the distribution of actions played by Alice against Bob is equal to the distribution of actions that results from the application of Alice's strategy to the distribution of signals that Alice observes when she is matched with Bob.[5]

We perturb the environment by introducing ϵ *committed agents* who each follow one strategy from an arbitrary finite set of *commitment strategies*.[6] We assume that at least one of the commitment strategies is totally mixed, which implies that all signals (i.e., all sequences of k actions) are observed with positive probability. A *steady state* in a perturbed environment describes a population in which $1-\epsilon$ of the agents are *normal* and follow strategies in S^* according to distribution σ^*, while ϵ of the agents follow commitment strategies. The behavior of normal and committed agents is described by the consistent behavior mapping η^*.

A steady state determines the long-run average payoff of all *incumbent strategies* (i.e., the normal strategies and the commitment strategies) in a straightforward way. Let $\pi(S^*, \sigma^*, \eta^*)$ denote the average payoff of the *normal agents* in the steady state (S^*, σ^*, η^*). A steady state also determines the payoff of a deviator (Alice) who deviates to a new strategy $\hat{s}$. Specifically, Alice's strategy determines her behavior against the incumbents. This determines the distribution of signals that are observed by the partners when being matched with Alice, and thus it determines the incumbents' play against Alice, and hence Alice's payoff, $\pi_{\hat{s}}(S^*, \sigma^*, \eta^*)$, in the new steady state that emerges following her deviation.

With these payoff definitions we are able to adapt the notions of Nash equilibrium, perfect equilibrium (Selten, 1975), and strict perfection (Okada, 1981) to our setup.[7] A steady state is a *Nash equilibrium* if no normal agent can gain by deviating to a different strategy, i.e., if $\pi_{\hat{s}}(S^*, \sigma^*, \eta^*) \leq \pi(S^*, \sigma^*, \eta^*)$ for any strategy $\hat{s}$. A steady state is a *perfect equilibrium* if it is the limit of a sequence of Nash equilibria in a converging sequence of perturbed environments. A pure action a^* is a *strictly perfect equilibrium action* if, for *any* converging sequence of perturbed environments, there is a converging sequence of Nash equilibria such that in the limit everyone plays a^*. That is, strict perfection requires stability with respect to all commitment strategies, whereas the stability of a perfect equilibrium may rely on the absence of some commitment strategies.

Summary of Results We begin our analysis with two results for general games. Our first result is that any Nash equilibrium of the underlying game can be implemented as a Nash equilibrium of the environment, for any value of k. Similarly, any perfect equilibrium of the underlying game can be implemented as a perfect equilibrium of the environment. Next, we demonstrate the usefulness of the refinement of strict perfection by showing that in coordination games only the Pareto-efficient Nash equilibrium satisfies strict perfection whenever agents observe at least two actions.

The remaining results of the paper focus on the Prisoner's Dilemma game, in which each player decides

[4] In the extension presented in Appendix A we allow agents to choose non-stationary strategies.

[5] The reason why the behavior mapping is required to be part of the description of a steady state, rather than being uniquely determined by the distribution of strategies, is that our environment, unlike a standard repeated game, lacks a global starting time that determines the initial conditions. An example of a distribution of strategies that has multiple consistent behaviors is as follows. The underlying game is the Prisoner's Dilemma, k is equal to three, and everyone plays the most frequently observed action in the sample of the three observed actions. There are three behaviors that are consistent with this population: one in which everyone cooperates, one in which everyone defects, and one in which everyone plays (on average) uniformly.

[6] In Section 6.3 we discuss how to extend our results to more general perturbed environments in which, in addition to commitment strategies, there are also observation errors or trembles.

[7] In Appendix B we show that our perfect equilibria also satisfy the refinement of evolutionary stability (Maynard-Smith, 1974).

simultaneously whether to cooperate or defect (see the payoff matrix in Table 1); if both players cooperate they obtain a payoff of one, if both defect they obtain a payoff of zero, and if one of the players defects, the defector gets $1+g$, while the cooperator gets $-l$, where $g, l > 0$ and $g < l+1$. (The latter inequality implies that mutual cooperation is the efficient outcome that maximizes the sum of payoffs.) We say that a Prisoner's Dilemma game is *offensive* if there is a stronger incentive to defect against a cooperator than against a defector (i.e., $g > l$); in a *defensive* Prisoner's Dilemma the opposite holds[8] (i.e., $g < l$). We start with a simple result (Prop. 4) that shows that defection is a strictly perfect equilibrium action for any number of observed actions.

Table 1: Matrix Payoffs of Prisoner's Dilemma Games

	c	d
c	1, 1	$-l$, $1+g$
d	$1+g$, $-l$	0, 0

Prisoner's Dilemma
G_{PD}: $g, l > 0$, $g < l+1$

	c	d
c	1, 1	-3, 2
d	2, -3	0, 0

Ex. 1: Defensive PD
G_D: $1 = g < l = 3$

	c	d
c	1, 1	-1.7, 3.3
d	3.3, -1.7	0, 0

Ex. 2: Offensive PD
G_O: $2.3 = g > l = 1.7$

Our first main result (Theorem 1) shows that always defecting is the unique perfect equilibrium in any offensive Prisoner's Dilemma game (i.e., $g > l$) for any number of observed actions. The result assumes a mild *regularity* condition on the set of commitment strategies (Def. 3), namely, that this set is rich enough such that, in any steady state of the perturbed environment, at least one of the commitment strategies induces agents to play a different distribution of actions than some of the normal agents.[9] The intuition is as follows. The mild assumption that not all agents defect with exactly the same probability, implies that the signal that Alice observes about her partner Bob is not completely uninformative. In particular, the more often Bob is observed to defect by Alice, the more likely it is that Bob will defect against Alice. In offensive games, it is better to defect against partners who are likely to cooperate, than to defect against partners who are likely to defect. This implies that a deviator who always defects is more likely to induce normal partners to cooperate. Consequently, such a deviator would outperform any agent who cooperates with positive probability.

Our anti-folk theorem result may come as a surprise in light of a number of existing papers that have presented various equilibrium constructions to support cooperation in any Prisoner's Dilemma game. As mentioned above (and discussed in more detail in Section 4.2), our result demonstrates that, in the presence of a small fraction of committed agents, mechanisms that have been proposed to support cooperation fail, regardless of how these committed agents play. In this way our paper provides a theoretical explanation of why experimental evidence suggests that subjects' behavior corresponds neither to contagious equilibria (see, e.g., Duffy and Ochs, 2009) nor to belief-free equilibria (see, e.g., Matsushima, Tanaka, and Toyama, 2013). Our result also shows that even when an agent observes many of her partner's actions, there is no way in which she can use this information to assess her partner's reputation, along the lines of the reputation mechanisms of Sugden (1986)

[8]This follows the terminology of Dixit (2003). Takahashi (2010) calls offensive (defensive) PDs submodular (supermodular). If cooperation is interpreted as exerting high effort, then the defensive Prisoner's Dilemma exhibits strategic complementarity: increasing one's effort from low to high is less costly if the opponent exerts high effort (as illustrated in Example 2 in Sect. 4.1).

[9]Propositions 1–2 study the implications of the mild regularity refinements in other games. Specifically, they show that (1) any perfect equilibrium of the underlying game that is not totally mixed can be implemented as a regular perfect equilibrium in any environment, and (2) the mild refinement rules out some totally mixed perfect equilibria of the underlying game, such as the totally mixed equilibrium in a coordination game.

and Kandori (1992, Theorem 2), which would otherwise ensure that cooperation can be sustained. Thus, despite their seemingly simple structure, the binary reputation mechanisms of Sugden (1986) and Kandori (1992, Theorem 2), cannot be implemented by plausible decentralized observation structures.[10]

Our second main result (Theorem 2) shows that cooperation is a strictly perfect equilibrium action in any defensive Prisoner's Dilemma game ($g < l$) when players observe at least two actions.[11] Moreover, there is an essentially unique distribution of strategies that support cooperation, according to which: (a) all agents cooperate when observing no defections (i.e., $m = 0$), (b) all agents defect when observing at least 2 defections ($m \geq 2$), (c) the normal agents defect with an average probability of $0 < q < 1$ when observing a single defection[12] ($m = 1$). The intuition for the result is as follows. Defection yields a direct gain that is increasing in the partner's probability of defection (due to the game being defensive). In addition, defection results in an indirect loss because it induces future partners to defect when they observe the current defection. This indirect loss is independent of the current partner's behavior. One can show that there always exists a probability q such that the above distribution of strategies balances the direct gain and the indirect loss of defection conditional on observing $m = 1$, while cooperation is the unique best reply when observing $m = 0$, and defection is the unique best reply when observing $m \geq 2$

Next, we analyze the case of observation of a single action (i.e., $k = 1$). Prop. 5 shows that cooperation is a perfect equilibrium action in a defensive Prisoner's Dilemma if and only if the bonus for defections is not too large (specifically, $g \leq 1$). The intuition is that similar arguments to the result above imply that there exists a unique average probability $q < 1$ by which agents defect when observing $m = 1$ in any cooperative perfect equilibrium. This implies that a deviator who always defects succeeds in getting a payoff of $1 + g$ in a fraction $1 - q > 0$ of the interactions, and that such a deviator outperforms the incumbents if g is too large.

Observations Based on Action Profiles So far we have assumed that each agent (Alice) observes only the partner's (Bob's) behavior against other opponents, but that she cannot observe the behavior of the past opponents against Bob. In Section 5 we relax this assumption. Specifically, we study three observation structures: the first two seem to be empirically relevant, and the third one is theoretically important since it allows us to construct an equilibrium that sustains cooperation in all Prisoner's Dilemma games.

1. *Observing conflicts*: each agent observes, in each of the k sampled interactions of her partner, 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 an observation structure (which we have not seen in the existing literature) seems like a plausible way to capture non-verifiable feedback about the partner's behavior.

2. *Observing action profiles:* each agent observes the full action profile in each of the sampled interactions.

3. *Observing actions against cooperation:* in each of the sampled interactions an agent observes what action the partner took provided that the partner's opponent cooperated. If the partner's opponent defected

[10] In a binary reputation mechanism each agent starts with a "good label" (or "good standing"). This label automatically becomes "bad" if a player defects against a "good" partner. The equilibrium strategy that supports full cooperation, given an exogenous mechanism that induces these labels, is to cooperate against "good" partners and defect against "bad" partners. See related models in Okuno-Fujiwara and Postlewaite (1995) and Ohtsuki and Iwasa (2006).

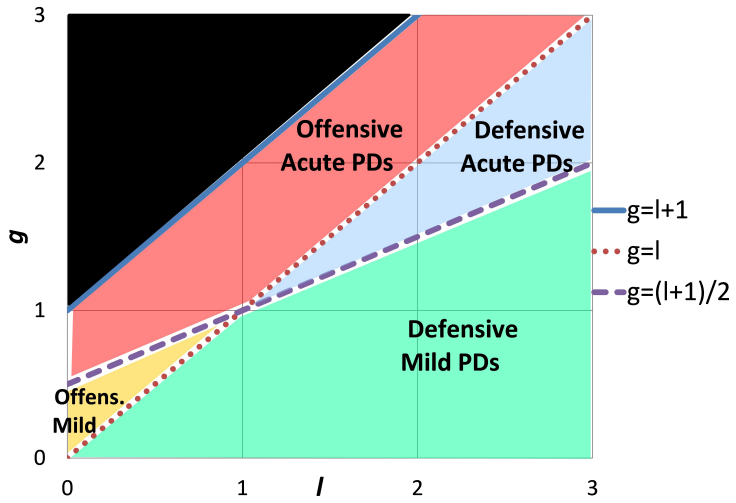
[11] Takahashi (2010) shows that there is a strict equilibrium that induces cooperation in defensive Prisoner's Dilemma games in environments in which each agent observes the *entire history* of past actions played by her partner. Theorem 2 shows that stable cooperation can be sustained in defensive Prisoner's Dilemma games also when players observe only two actions.

[12] The specific commitment strategies that are present in the perturbed environment influence two aspects in the perfect equilibrium that supports cooperation: (1) they affect the average defection probability when an agent observes a single defection, and (2) they determine whether each agent mixes when she observes a single defection or whether the population includes two different groups of agents, such that only agents in one of these groups defect when they observe a single defection.

then there is no information about what the partner did.

It turns out that the stability of cooperation in the first two observation structures crucially depends on a novel classification of the Prisoner's Dilemma games. We say that a Prisoner's Dilemma game is *acute* if $g > \frac{l+1}{2}$, and *mild* if $g < \frac{l+1}{2}$. The threshold between the two categories, namely $g = \frac{l+1}{2}$, 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. Consider a setup 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 Prisoner's Dilemmas requires this probability to be more than 50%, while a probability of below 50% is enough to deter deviations for mild PDs. Figure 1 illustrates the classification of games into mild/acute/offensive/defensive (see Section 5.2 for further discussion).

Figure 1: Classification of Prisoner's Dilemma Games



Our next results (Theorems 3–4) show that in both observation structures (conflicts or action profiles, and assuming $k \geq 2$) cooperation is a perfect equilibrium action if and only if the underlying Prisoner's Dilemma game is mild. Moreover, cooperation is supported by essentially the same unique behavior as in Theorem 2. However, the mixing probability q is determined in a somewhat different way. The reason why cooperation cannot be sustained in acute games with observation of conflicts is as follows. Suppose an agent considers defecting with probability γ. The direct gain when facing a cooperator is then $\gamma \cdot g$. If γ is small then the probability that a future opponent observes the agent to have been involved in a conflict in at least one out of k interactions is roughly $\gamma \cdot k$. Thus, in order to deter agents from defecting, it must be that the normal agents' average probability of defection (q) when observing a single conflict satisfies $\gamma \cdot k \cdot q \cdot (l+1) > \gamma \cdot g$, or equivalently $k \cdot q \cdot > \frac{2g}{(l+1)} > \frac{1}{2}$. However, such a high value of $k \cdot q$ implies that defection is contagious: each defection results in a conflict being registered for both players and when a future partner observes a conflict he defects with a probability of more than 50%. Thus the fraction of defections grows steadily, until all normal agents defect with a high probability.

The intuition for why cooperation cannot be sustained in acute games with observation of action profiles is as follows. The fact that $k \cdot q$ must be larger than 50% in order to deter defections in acute games implies that when an agent (Alice) observes her partner (Bob) to defect against a cooperative opponent, then Bob is more likely to do so because he is a normal agent who observed his past opponent to defect, than because Bob is a

committed agent. This implies that Alice puts a higher probability on Bob defecting against her conditional on her observing Bob to have defected against a partner who also defected, than she does conditional on her observing Bob to have defected against an opponent who cooperated. Thus, defecting is the unique best reply when observing the partner defect against a defector, but it removes the incentives required to support stable cooperation.

Finally, we show that the third observation structure, *observing actions against cooperation*, is optimal in the sense that it sustains cooperation as a perfect equilibrium action for any Prisoner's Dilemma game (Theorem 5). In this environment, each agent (Alice) observes Bob's action only in interactions in which Bob's opponent cooperated. Specifically, the signal in each sampled interaction has three possible values: either both players cooperated, or Bob unilaterally defected, or Bob's partner defected (and then Alice cannot observe Bob's action). Not allowing Alice to observe Bob's behavior against a defector helps to sustain cooperation because it implies that defecting against a defector does not have any negative indirect effect (in any steady state) because it is never observed by future opponents. This encourages agents to defect against partners who are more likely to defect (regardless of the values of g and l).

The following table summarizes most of the results about the stability of cooperation.

Table 2: Stability of Cooperation in the Prisoner's Dilemma

Category of PD	Parameters	Observation Structure (2+ Sampled Interactions)			
		Actions	Conflicts	Action profiles	actions against cooperation
Mild & Defensive	$g < \min\left(l, \frac{l+1}{2}\right)$	**Y**	**Y**	**Y**	**Y**
Mild & Offensive	$l < g < \frac{l+1}{2}$	**N**			
Acute & Defensive	$\frac{l+1}{2} < g < l$	**Y**	**N**	**N**	
Acute & Offensive	$\max\left(l, \frac{l+1}{2}\right) < g$	**N**			

Empirical Predictions Our paper yields various novel testable empirical predictions in a setup in which agents are randomly matched to play the Prisoner's Dilemma and each agent observes some information about the partner's past interactions. Three of these testable predictions are as follows. First, we predict that when players observe the partner's past actions, then cooperation can be sustained only in defensive Prisoner's Dilemma games. Second, we predict then when players observe past conflicts or past action profiles of the partner, then cooperation can be sustained only in mild Prisoner's Dilemma games. Finally, we predict that players are more likely to defect, the more times they observe the partner to be involved in defections.

Currently, there is only some experimental evidence regarding behavior in the kind of setups that we model. The existing evidence is mainly relevant to, and supportive of, our last prediction (see, e.g., Wedekind and Milinski, 2000; Milinski, Semmann, Bakker, and Krambeck, 2001; Seinen and Schram, 2006; Engelmann and Fischbacher, 2009). In Section 6.2 we discuss our empirical predictions in detail, describe existing experimental findings, and relate our predicted comparative statics to those obtained in papers studying repeated games between fixed pairs of players (see, e.g., Blonski, Ockenfels, and Spagnolo, 2011; Dal Bó and Fréchette, 2011; Breitmoser, 2015).

Related Literature A substantial literature studies the possibility of sustaining stable cooperation when agents from a large population are randomly matched to play the Prisoner's Dilemma game (see, e.g., Nowak and Sigmund's (2005) survey on indirect reciprocity). In what follows we discuss the contribution of our paper with respect to this literature, excluding the literature discussed earlier in the Introduction.

A few papers (e.g., van Veelen, García, Rand, and Nowak, 2012; Alger and Weibull, 2013) show that it is possible to sustain cooperation with no information about the partner's behavior if matching is sufficiently assortative; i.e., cooperators are more likely to interact with other cooperators.[13] Our paper shows that letting players observe the partner's behavior in two interactions is sufficient to sustain cooperation without assuming assortativity.

In an influential paper, Nowak and Sigmund (1998) presents the mechanism of *image scoring* to support cooperation when players observe the partner's past actions. In their setup, each agent observes the last k past actions of the partner, and she defects if and only if the partner has defected at least m times in the last k observed actions. A couple of papers have raised concerns about the stability of cooperation under image-scoring mechanisms. Specifically, Leimar and Hammerstein (2001) demonstrate in simulations that cooperation is unstable, and Panchanathan and Boyd (2003) analytically study the case in which each agent observes the last action.[14] Our paper makes three key contributions with respect this literature. First, we introduce a novel variant of image scoring that is essentially the unique way to support stable cooperation when observing actions. Second, we show that cooperation can be supported by image-scoring mechanisms only if an agent faces some uncertainty about the what the partner has observed about her (more precisely, the observed k actions have to be sampled from a larger history and to be privately observed by the partner; see the analysis in Appendix A). Third, we show that the classification of Prisoner's Dilemma games into offensive and defensive games is critical to the stability of cooperation when agents observe actions. We demonstrate that popular existing versions of image scoring do not support cooperation in offensive Dilemma games.

Takahashi (2010) presents a related result with a similar classification of Prisoner's Dilemma games. He shows that (1) always defecting is the unique strict equilibrium in offensive Prisoner's Dilemma games, and (2) there is a strict equilibrium that induces full cooperation in defensive games; this equilibrium relies on the agents observing many past actions and on following a complex strategy.[15] Our result substantially enhances Takahashi's result in at least two ways. First, we show the anti-folk theorem result for offensive games while using a much weaker solution concept, namely, a perfect equilibrium rather than a strict equilibrium. Second, we support stable cooperation in defensive Prisoner's Dilemma games in setups in which agents observe only two actions, and use very simple strategies. Theorem 1 shows that the mild requirement of robustness to the inclusion of a few committed agents destabilizes all perfect equilibria except the one in which everyone always defects. Bhaskar (1998) presents a related result, though he deals with a very different setup and refinement. Specifically, Bhaskar studies the sustainability of inter-generational transfers in Samuelson's consumption-loan model, and shows that only the "defective" no-transfers equilibrium survives the mild requirements of Harsanyi's purification and a bit of uncertainty about actions played in the distant past.

Structure Section 2 presents the model. Our solution concept is described in Section 3. Section 4 contains our main results. Section 5 extends the model to deal with general observation structures. In Section 6 we

[13] See also the following papers that study the stability of cooperation in other kinds of structured populations: Herold (2012) who studies a "haystack" model in which individuals interact within separate groups, Fujiwara-Greve and Okuno-Fujiwara (2009) who study a "voluntarily separable" repeated Prisoner's Dilemma; and Cooper and Wallace (2004) who study group selection.

[14] See Berger and Grüne (2014) who study observation of k actions, but restrict agents to play only image-scoring-like strategies.

[15] The strategy in Takahashi (2010) treats the entire repeated game as if it were T separate subgames (those occurring in rounds 1 modulo T, those occurring in rounds 2 modulo T, etc.), and it induces players to play a grim-trigger strategy in each subgame. Note, that agents need to know the calendar time perfectly, and that $T \to \infty$ when the players' discount factor converges to one.

discuss our empirical predictions, the robustness of our results, and the introduction of cheap talk. Section 7 concludes. Additional results and detailed proofs are contained in the online appendices. Appendix A extends our model to deal with finitely lived agents, non-stationary strategies, and observing the most recent actions of the partner. Appendix B presents the refinement of evolutionary stability. The formal proofs appear in Appendix C.

2 Model

2.1 Environment

We model an environment in which patient agents in a large population are randomly matched at each round to play a two-player symmetric one-shot game. For tractability we assume throughout the paper that the population is a continuum.[16] In the main model we further assume that the agents are infinitely lived and do not discount the future (i.e., they maximize the average per-round long-run payoff). Alternatively, our main model can be interpreted as representing interactions between finitely lived agents who belong to infinitely lived dynasties, such that an agent who dies is succeeded by a protégé who plays the same strategy as the deceased mentor, and each agent observes k random actions played by the partner's dynasty. In Appendix A we show how to extend our results to a setup with finitely lived agents, in which new agents begin with a "blank history."

Before playing the game, each agent (she) privately observes k random actions that her partner (he) played against other opponents in the past.[17] As described in detail below, in the main model agents are restricted to use only stationary strategies, such that the agent's behavior depends only on the signal about the partner, and not on the agent's own past play or on time. (This assumption is relaxed in Appendix A.) Thus each agent's behavior results in a well-defined aggregate stationary distribution of actions. The k actions that an agent observes about her partner are drawn independently from the partner's stationary distribution of actions. This sampling procedure may be interpreted as the limit of a process in which each agent randomly observes k actions that are uniformly sampled from the last n interactions of the partner, and one lets $n \to \infty$. (A formal related construction is presented in Appendix A.)

An environment is a pair $E = (G, k)$, where $G = (A, \pi)$ is a two-player symmetric normal-form game, and $k \in \mathbb{N}$ is the number of observed actions. Let $A = \{a_1, ..., a_{|A|}\}$ be the finite set of actions, and let $\pi : A \times A \to \mathbb{R}$ be the payoff function of the underlying game. Let $\Delta(A)$ denote the set of mixed actions (distributions over A), and let π be extended to mixed actions in the usual linear way. We use the letter a (respectively α) to denote a typical pure (respectively mixed) action. With a slight abuse of notation let $a \in A$ also denote the element in $\Delta(A)$ that assigns probability 1 to a. *We adopt this convention for all probability distributions throughout the paper.*

Remark 1. The assumption that the underlying 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 the agent's strategy conditions his played action on the assigned role (see, e.g., Selten, 1980).

[16] The results can be adapted to a setup with a large finite population. We do not formalize a large finite population, as this adds much complexity to the model without giving substantial new insights. Most of the existing literature also models large populations as a continuum (see, e.g., Rubinstein and Wolinsky 1985; Weibull 1995; Dixit 2003; Dekel, Ely, and Yilankaya 2007; Alger and Weibull 2013). Kandori (1992) and Ellison (1994) show that large finite populations differ from infinite populations because only the former can induce contagious equilibria. However, as noted by Ellison (1994, p. 578), and as discussed in Section 4.2, these contagious equilibria fail in the presence of a single "crazy" agent who always defects (also in a finite population).

[17] We restrict attention to a fixed number of observed actions to simplify the presentation and the notation. As discussed in Comment 4 in Section 4.3, our results can be extended to a setup in which the number of observed actions is random.

2.2 Stationary Strategy

The signal observed about the partner is the number of times he played each action $a \in A$ in the sample of k observed actions. Fix an environment $E = ((A, \pi), k)$. Let M denote the set of feasible signals:

$$M = \left\{ m \in \mathbb{N}^{|A|} \middle| \sum_i m_i = k \right\},$$

where m_i is interpreted as the number of times that action a_i has been observed in the sample.

Given a distribution of actions $\alpha \in \Delta(A)$ and an environment $E = ((A, \pi), k)$, let $\nu_\alpha(m_1, ..., m_{|A|})$ be the probability of observing signal $(m_1, ..., m_{|A|})$ conditional on being matched with a partner who plays on average the distribution of actions $\alpha \in \Delta(A)$. That is, $\nu_\alpha \in \Delta(M)$ is a multinomial distribution that describes a sample of k i.i.d. actions, where each action is distributed according to α:

$$\forall (m_1, ..., m_{|A|}) \in M, \quad \nu_\alpha(m_1, ..., m_{|A|}) = \left(\frac{k!}{m_1! \cdots m_{|A|}!} \cdot (\alpha(a_1))^{m_1} \cdots (\alpha(a_{|A|}))^{m_{|A|}} \right). \tag{1}$$

A *stationary strategy* (henceforth, *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 assigned 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. Strategy s is *totally mixed* if for each action $a \in A$ and signal $m \in M$ $s_m(a) > 0$. Let $\mathcal{S}$ denote the set of all strategies. Given strategy s and distribution of signals $\nu \in \Delta(M)$, let $s_\nu \in \Delta(A)$ be the distribution of actions played by an agent who follows strategy s and observes a signal sampled from ν:

$$\forall a \in A, \quad s_\nu(a) = \sum_{m \in M} \nu(m) \cdot s_m(a).$$

2.3 Steady State

A steady state of an environment (G, k) is a triple: (1) a finite set of strategies S^* interpreted as the strategies that are played by the agents in the population, (2) a distribution σ over S^* interpreted as describing the fraction of agents following each strategy, and (3) a function $\eta_s(s')$ that describes the distribution of actions played by each agent who follows strategy s conditional on being matched against an agent who follows strategy s'; this function is required to describe a consistent behavior in the sense that the distribution of actions played by Alice against Bob is the one that is induced by the distribution of signals that Alice observes when being matched with Bob (which is determined by the average distribution of actions played by Bob against a random partner), together with Alice's strategy. Formally:

Definition 1. A *steady state* (or *state* for short) of an environment (G, k) is a triple (S, σ, η) where $S \subseteq \mathcal{S}$ is a finite set of strategies, $\sigma \in \Delta(S)$ is a distribution with full support over S, and $\eta : S \times S \to \triangle(A)$ is a mapping (called, *consistent behavior*) that assigns to each pair of strategies $s, s' \in S$ a mixed action $\eta_s(s') \in \Delta(A)$ and satisfies the consistency condition below. If $\eta_s(s') = a$ for each s, s', then we denote it by $\eta \equiv a$. Let $\bar{\eta}_s$ be the average distribution of actions played by an agent who follows strategy s against a random partner:

$$\forall a \in A, \quad \bar{\eta}_s(a) := \sum_{s' \in S} \sigma(s') \cdot \eta_s(s')(a). \tag{2}$$

The *consistency requirement* that the mapping η has to satisfy is:

$$\forall a \in A,\ s, s' \in S,\quad \eta_s\left(s'\right)(a) = s_{\nu_{\bar{\eta}_{s'}}}(a). \tag{3}$$

The consistency requirement (3) should be interpreted as follows. When Alice (who follows strategy s) is being matched with Bob (who follows strategy s'), she observes each signal m with probability $\nu_{\bar{\eta}_{s'}}(m)$ because $\nu_{\bar{\eta}_{s'}}$ is the distribution of signals induced by Bob's distribution of actions against a random partner (whose actions are distributed according to $\bar{\eta}_{s'}$). Conditional on observing signal m, Alice plays each action a with probability $s_m(a)$.

A standard fixed-point argument shows that any distribution of strategies admits a consistent behavior.

Lemma 1. *Let S be a finite set of strategies and let $\sigma \in \Delta(S)$ be a distribution. Then, there exists a mapping $\eta : S \times S \to \Delta(A)$ such that (S, σ, η) is a steady state.*

Some distributions induce multiple consistent behaviors. For example, if the underlying game is the Prisoner's Dilemma, each agent observes 3 of the partner's actions (i.e., $k = 3$), and everyone follows the strategy of playing the most frequently observed action (i.e., with the terminology introduced below $S = \left\{s^2\right\}$, where $s^2(m) = d$ iff $m \geq 2$), then there are three consistent behaviors: one in which everyone cooperates (i.e., $\eta_{s^2}\left(s^2\right)(d) = 0$), one in which everyone defects (i.e., $\eta_{s^2}\left(s^2\right)(d) = 1$), and one in which everyone plays (on average) uniformly[18] (i.e., $\eta_{s^2}\left(s^2\right)(d) = 0.5$).

2.4 Perturbed Environment

In a seminal paper Kreps, Milgrom, Roberts, and Wilson (1982) show, in a standard setup of a two-player finitely repeated Prisoner's Dilemma, that the equilibrium analysis completely changes if one slightly perturbs the environment by assuming that with very small probability one of the players may be committed to following a "tit-for-tat" strategy. (See Mailath and Samuelson, 2006, for a textbook analysis and a survey of the "reputation" literature.) Motivated by this observation, we introduce a notion of perturbed environments in which a small fraction of agents in the population are committed to playing specific strategies, even though these strategies are not necessarily payoff-maximizing.

A perturbed environment is a tuple consisting of: (1) an environment, (2) a distribution λ over a set of commitment strategies S_C that includes a totally mixed strategy, and (3) a number ϵ representing how many agents are committed to playing strategies in S_C (henceforth, *committed agents*). The remaining $1 - \epsilon$ of the agents can play any strategy in $\mathcal{S}$ (henceforth, *normal agents*). The set of commitment strategies and the set of normal strategies together constitute the set of incumbent strategies $S \cup S_C$. Formally:

Definition 2. A *perturbed environment* is a tuple $E_\epsilon = ((G, k), (S_C, \lambda), \epsilon)$, where G is the underlying game, $k \in \mathbb{N}$ is the number of observed actions, S_C is a non-empty finite set of strategies (called, *commitment strategies*) that includes a totally mixed strategy, $\lambda \in \Delta(S_C)$ is a distribution with full support over the commitment strategies, and $\epsilon \geq 0$ is the mass of committed agents in the population.

We require a S_C to include at least one totally mixed strategy because we want all signals to be observed with a positive probability in a perturbed environment when $\epsilon > 0$ (analogous to the requirement in Selten (1975) that all actions be played with a positive probability in the perturbations defining a perfect equilibrium).

[18] In Heller and Mohlin (2015b) we study a setup in which the number of observed actions is random, and we show that all strategy distributions admit unique consistent behaviors iff the expected number of observed actions is less than one.

We refer to (S_C, λ) as a *distribution of commitments*. Let 1_s denote the degenerate distribution that assigns mass 1 to strategy s. With a slight abuse of notation, we identify an *unperturbed environment* (with $\epsilon = 0$) $((G,k),(S_C,\lambda),\epsilon=0)$ with the equivalent environment (G,k).

Remark 2. To simplify the presentation, the definition of perturbed environment includes only commitment strategies, and it does not allow "trembling hand" mistakes. As discussed in Section 6.3, the results also hold in a setup in which agents also tremble, as long as the probability by which a normal agent trembles is of the same order of magnitude as the frequency of committed agents.

Our anti-folk theorem result (Theorem 1) requires an additional mild assumption on the perturbed environment that rules out the knife-edge case in which all agents (committed and non-committed alike) behave exactly the same. Specifically, a set of commitments is regular if for each distribution of actions α, there exists a committed strategy s that does not play distribution α when observing the signal distribution induced by α. Formally:

Definition 3. A set of commitment strategies S_C is *regular* if for each distribution of actions $\alpha \in \Delta(A)$, there exists a strategy $s \in S_C$ such that $s_{\nu_\alpha} \neq \alpha$.

If the set of commitments is regular, then we say that the distribution (S_C, λ) and the perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ are regular. An example of a regular set of commitments is the set that includes two strategies $s \equiv \alpha_1$ and $s' \equiv \alpha_2$ that induce agents to play mixed actions $\alpha_1 \neq \alpha_2$ regardless of the observed signal.

2.5 Steady State in a Perturbed Environment

A steady state (S,σ,η) of a perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ is defined in a similar way to Def. 1, with the following adaptations: (1) the finite set of strategies S is interpreted as the strategies followed by the normal agents, and (2) the function $\eta_s(s')$ describes the distribution of actions played by *incumbent agents* (either normal or committed agents), and it has to satisfy an analogous consistency requirement as in Def. 1. Formally:

Definition 4. A *steady state* (or *state*) of a perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ is a triple (S,σ,η) where $S \subseteq \mathcal{S}$ is a finite set, $\sigma \in \Delta(S)$ is a distribution with a full support over S, and $\eta : (S \cup S_C) \times (S \cup S_C) \to \triangle(A)$ is a mapping (called, *consistent behavior*) that assigns to each pair of strategies $s, s' \in S \cup S_C$ a mixed action $\eta_s(s') \in \Delta(A)$ and satisfies the consistency condition below. Let $\bar{\eta}_s$ be the average distribution of actions played by an agent who follows strategy s against a random partner:

$$\forall a \in A, \quad \bar{\eta}_s(a) := \sum_{s' \in S \cup S_C} \left((1-\epsilon)\cdot\sigma(s') + \epsilon\cdot\lambda(s')\right)\cdot\eta_s(s')(a). \tag{4}$$

The *consistency requirement* that the mapping η has to satisfy is:

$$\forall a \in A,\ s, s' \in S \cup S_C, \quad \eta_s(s')(a) = s_{\nu_{\bar{\eta}_{s'}}}(a). \tag{5}$$

Note that the definitions of perturbed environment and steady state imply that $\sigma(s) + \lambda(s) > 0$ if and only if s is an incumbent strategy.

The following example demonstrates a specific steady state in a specific perturbed environment in the Prisoner's Dilemma game. The example is intended to clarify the various definitions of this section, and, in particular, to clarify the consistency requirement. Later, we revisit to the same example to demonstrate the essentially unique perfect equilibrium that supports stable cooperation.

Example 1. Consider the perturbed environment $\left(\left(G_{PD},2\right),\left(\left\{s^u\equiv 0.5\right\},1_{s^u}\right),\epsilon\right)$, in which the underlying game is the Prisoner's Dilemma, each agent observes two of her partner's actions, there is a single commitment strategy, denoted s^u, which is followed by a fraction $0<\epsilon<<1$ of committed agents, who choose each action with probability 0.5 regardless of the observed signal. Let $\left(S^*=\left\{s^1,s^2\right\},\sigma^*=\left(\frac{1}{6},\frac{5}{6}\right),\eta^*\right)$ be the following steady state. The state includes two normal strategies: s^1 and s^2. The strategy s^1 defects iff $m\geq 1$, and the strategy s^2 defects iff $m\geq 2$. The distribution σ assigns a mass of $\frac{1}{6}$ to s^1 and a mass of $\frac{5}{6}$ to s^2. The probability of defection in the consistent behavior η^* is defined as follows (neglecting terms of $O\left(\epsilon^2\right)$):

$$\eta^*_{s^u}\left(\cdot\right)\left(d\right)=50\%,\ \ \eta^*_{s^1}\left(s'\right)\left(d\right)=\begin{cases}75\% & s'=s^u\\ 3.5\cdot\epsilon+O\left(\epsilon^2\right) & s'=s^1\\ 0.5\cdot\epsilon+O\left(\epsilon^2\right) & s'=s^2,\end{cases}\quad \eta^*_{s^1}\left(s'\right)\left(d\right)=\begin{cases}25\% & s'=s^u\\ O\left(\epsilon^2\right) & s'=s^1\\ O\left(\epsilon^2\right) & s'=s^2.\end{cases} \tag{6}$$

The average probability of defection for each strategy is given by:

$$\bar{\eta}^*_{s^u}=50\%,\ \ \bar{\eta}_{s^1}=1.75\cdot\epsilon+O\left(\epsilon^2\right),\ \ \bar{\eta}^*_{s^2}=0.25\cdot\epsilon+O\left(\epsilon^2\right),$$

where $\bar{\eta}^*_{s^1}$ (for example) is calculated as follows:

$$\bar{\eta}^*_{s^1}=\sum_{s'\in S\cup S_C}\sigma\left(s'\right)\cdot\eta_{s^1}\left(s'\right)=\epsilon\cdot 75\%+\frac{1-\epsilon}{6}\cdot 3.5\cdot\epsilon+\frac{5\cdot\left(1-\epsilon\right)}{6}\cdot 0.5\cdot\epsilon=1.75\cdot\epsilon+O\left(\epsilon^2\right).$$

This implies that the distributions of signals induced by each strategy are the following binomial distributions:

$$\nu_{\bar{\eta}^*_{s^u}}\sim Bin\left(2,50\%\right),\ \ \nu_{\bar{\eta}^*_{s^u}}\left(0\right)=25\%,\ \nu_{\bar{\eta}^*_{s^u}}\left(1\right)=50\%,\ \nu_{\bar{\eta}_{s^u}}\left(2\right)=25\%,$$

$$\nu_{\bar{\eta}^*_{s^1}}\sim Bin\left(2,1.75\cdot\epsilon+O\left(\epsilon^2\right)\right),\ \ \nu_{\bar{\eta}^*_{s^1}}\left(0\right)=100\%-O\left(\epsilon\right),\ \nu_{\bar{\eta}^*_{s^1}}\left(1\right)=3.5\cdot\epsilon+O\left(\epsilon^2\right),\ \nu_{\bar{\eta}^*_{s^1}}\left(2\right)=O\left(\epsilon^2\right),$$

$$\nu_{\bar{\eta}^*_{s^2}}\sim Bin\left(2,0.25\cdot\epsilon+O\left(\epsilon^2\right)\right),\ \ \nu_{\bar{\eta}^*_{s^2}}\left(0\right)=100\%-O\left(\epsilon\right),\ \nu_{\bar{\eta}^*_{s^2}}\left(1\right)=0.5\cdot\epsilon+O\left(\epsilon^2\right),\ \nu_{\bar{\eta}^*_{s^2}}\left(2\right)=O\left(\epsilon^2\right).$$

One can verify that indeed the consistency requirement is satisfied: $\eta^*_s\left(s'\right)=s_{\nu_{\bar{\eta}_{s'}}}$ for each pair of strategies $s,s'\in\left\{s^1,s^2,s^u\right\}$. For example, strategy s^1 inducing defection with probability $0.5\cdot\epsilon+O\left(\epsilon^2\right)$ when being matched with strategy s^2 (i.e., $\eta^*_{s^1}\left(s^2\right)=0.5\cdot\epsilon+O\left(\epsilon^2\right)$ is consistent with the distribution of signals induced by an s^2-agent, which yields $m\geq 1$ with probability of $0.5\cdot\epsilon+O\left(\epsilon^2\right)$.

2.6 Discussion of the Model

Most of the existing literature represents interactions within a community as a repeated game (e.g., Kandori, 1992; Ellison, 1994; Dixit, 2003; Takahashi, 2010; Deb, 2012; Deb and González-Díaz, 2014). A repeated game (or more generally, an extensive-form game) has a "global time zero," in which the first ever interaction takes place. In many real-life situations, the interactions within a community began a long time ago and have continued, via overlapping generations, to the present day. It seems implausible that today's agents condition their behavior on what happened in the remote past (or on calendar time). For example, trade interactions have been been taking place from time immemorial. It seems unreasonable to assume that Alice's behavior today is conditioned on what transpired in some long-forgotten time $t=0$, when, say, two hunter-gatherers were involved in the first ever trade. We suggest that, even though real-world interactions obviously begin at some definite date, the best way of modeling how the interacting agents think about the situation themselves

may be to get rid of global time zero and focus on strategies that do not condition on what happened in the remote past, or on calendar time.[19] Consequently our concept of an environment differs from a repeated game in that it lacks a global time zero. This is the reason why unlike in repeated games, a distribution of strategies does not uniquely determine the behavior and the payoffs of the agent, so that one must explicitly add the consistent behavior η as part of the description of the state of the population.

A few papers follow an approach that is similar to ours in that they study steady states of interactions within a community and do not assume a global time zero. Rosenthal (1979) develops the notion of a steady-state Nash equilibrium in environments in which each player observes the partner's last action, and applies it to study the Prisoner's Dilemma. Rosenthal focuses only on pure steady states (in which everyone uses the same pure strategy), and concludes that defection is the unique pure stationary Nash equilibrium action except in a few knife-edge cases. Other papers following a related approach include Rubinstein and Wolinsky (1985), who study bargaining, Young (1993), who studies the evolution of conventions,[20] Dekel, Ely, and Yilankaya (2007), who study non-material preferences, and Eliaz and Rubinstein (2014), who study boundedly rational agents. Acemoglu and Wolitzky (2014) include a global time zero but still do not allow players to condition their behavior on the full history, or on calendar time. Instead they assume that players at $t \geq 1$ have an "improper uniform prior" about the calendar time.

Our model is novel in at least two key respects in relation to the previous literature: (1) we allow each agent to observe the behavior of the partner in several past interactions (and in Section 5 we also allow her to observe the behavior directed against the partner in the past), and (2) we introduce a small fraction of committed agents and adapt the notion of a perfect equilibrium to this setup (see, Section 3).

We think of our notion of a steady state as capturing plausible stable long-run outcomes of some (unmodeled) dynamic adjustment process by which agents sometimes experiment and gradually revise their strategy choices in response to how well different strategies perform. Relatedly, it is possible to interpret a steady state (S, σ, η) as a kind of initial condition for society, in which agents already have a long-existing past. That is, we begin our analysis of community interaction at a point in time when agents have already followed the strategy distribution (S, σ) and behaved according to the consistent behavior η for a very long time. We then ask whether the agents have any profitable deviation from their strategy. If not, then the steady state (S, σ, η) is likely to continue to persist. This approach stands in contrast with the standard approach that studies whether or not agents have a profitable deviation at time $t >> 1$ following a long history that started with the first ever interaction at $t = 0$.

3 Solution Concept

3.1 Long-Run Payoff

In this subsection we define the long-run average (per-round) payoff of a patient agent who follows a stationary strategy s, given a steady state (S, σ, η) of a perturbed environment $((G, k), (S_C, \lambda), \epsilon)$. The same definition, when taking $\epsilon = 0$, holds for an unperturbed environment.

We define the long-run payoff of an agent who follows the incumbent strategy $s \in S \cup S_C$ as:

$$\pi_s(S, \sigma, \eta) = \sum_{s' \in S \cup S_C} \left((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')\right) \cdot \left(\sum_{(a,a') \in A \times A} \eta_s(s')(a) \cdot \eta_{s'}(s)(a') \cdot \pi(a, a') \right). \quad (7)$$

[19] For a related discussion of the proper way of modeling finitely and infinitely repeated interactions, see Osborne and Rubinstein (1994, p. 135).

[20] Young (1993) differs from the other papers by having an explicit long-run dynamics that selects one of the steady states regardless of the initial conditions.

Eq. (7) is straightforward. The inner (right-hand) sum (i.e., $\sum_{(a,a')\in A\times A}\eta_s(s')(a)\cdot\eta_{s'}(s)(a')\cdot\pi(a,a')$) calculates the expected payoff of Alice who follows strategy s conditional on being matched with a partner who follows strategy s' . The outer sum weighs these conditional expected payoffs according to the frequency of each incumbent strategy s' (i.e., $((1-\epsilon)\cdot\sigma(s')+\epsilon\cdot\lambda(s'))$), which yields the expected payoff of Alice against a random partner in the population.

Let $\pi(S,\sigma,\eta)$ be the average payoff of the *normal* agents:

$$\pi(S,\sigma,\eta)=\sum_{s\in S}\sigma(s)\cdot\pi_s(S,\sigma,\eta).$$

Now consider a normal agent (Alice) who deviates and plays a new strategy $\hat{s}\in\mathcal{S}\backslash S$. If Alice deviates to a commitment strategy $\hat{s}\in S_C$, then her payoff is simply $\pi_{\hat{s}}(S,\sigma,\eta)$, as defined in Eq. (7). Suppose Alice deviates to a non-incumbent strategy $\hat{s}\in\mathcal{S}\backslash(S\cup S_C)$. Alice's strategy determines her behavior against the incumbents. This determines the distribution of signals that are observed by the partners when being matched with Alice, and thus determines the incumbents' play against Alice, and Alice's payoff $\pi_{\hat{s}}(S,\sigma,\eta)$ in the new steady state that emerges following her deviation. Formally:

Definition 5. Given steady state (S,σ,η) in environment $((G,k),(S_C,\lambda),\epsilon)$ and a non-incumbent strategy $\hat{s}\in\mathcal{S}\backslash(S\cup S_C)$, let $(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta})$ be the *post-deviation steady state* that is reached when a single deviator plays strategy $\hat{s}$:

1. $\hat{\sigma}(s)=\sigma(s)$ for each $s\in S$ and $\hat{\sigma}(\hat{s})=0$.
2. $\hat{\eta}_s(s')=\eta_s(s')$ for each $s,s'\in S\cup S_C$.
3. $\forall a\in A$, $s'\in S\cup S_C$ $\hat{\eta}_{\hat{s}}(s')(a)=\hat{s}_{\nu_{\bar{\eta}_{s'}}}(a)$.
4. $\forall a\in A$, $s\in S\cup S_C$ $\hat{\eta}_s(\hat{s})(a)=s_{\nu_{\bar{\hat{\eta}}_{\hat{s}}}}(a)$, where $\bar{\hat{\eta}}_{\hat{s}}$ denotes the average distribution of actions of the deviator (i.e., for each $a\in A$, $\bar{\hat{\eta}}_{\hat{s}}(a)=\sum_{s'\in S\cup S_C}((1-\epsilon)\cdot\sigma(s')+\epsilon\cdot\lambda(s'))\cdot\hat{\eta}_{\hat{s}}(s')(a)$).

The interpretation of the various parts of Def. 5 is as follows. Part (1) reflects the fact that Alice (the agent who deviates to $\hat{s}$) has mass zero in the infinite population. This also implies that Alice's deviation does not affect the behavior of the two incumbents who are matched to play together (which is reflected in Part (2)). Part (3) states that Alice's behavior against an incumbent Bob, who plays s', is the behavior that is induced by her strategy together with the distribution of signals she observes about Bob. Finally, Part (4) states that the behavior of an incumbent, Bob, against Alice, is the behavior that is induced by his strategy together with the distribution of signals he observes about Alice.

Let $\pi_{\hat{s}}(S,\sigma,\eta)$ denote the payoff of someone who deviates to a $\hat{s}\in\mathcal{S}\backslash(S\cup S_C)$, i.e., the payoff to $\hat{s}$ in the post-deviation steady state $(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta})$:

$$\pi_{\hat{s}}(S,\sigma,\eta):=\pi_{\hat{s}}(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta}).$$

3.2 Nash and Perfect Equilibrium

A steady state is a Nash equilibrium if no agent can obtain a higher payoff by a unilateral deviation. Formally:

Definition 6. The steady state (S,σ,η) of perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ is a *Nash equilibrium* if for each strategy $s\in\mathcal{S}$: $\pi_s(S,\sigma,\eta)\leq\pi(S,\sigma,\eta)$.

Note that the $1-\epsilon$ normal agents in such a Nash equilibrium must obtain the same maximal payoff. That is, each normal strategy $s \in supp(\sigma)$ satisfies $\pi_s(S,\sigma,\eta) = \pi(S,\sigma,\eta) \geq \pi_{s'}(S,\sigma,\eta)$ for each strategy $s' \in \mathcal{S}$. However, the ϵ committed agents may obtain lower payoffs.

Next, observe that any symmetric Nash equilibrium (α,α) of the underlying game can be implemented in a corresponding Nash equilibrium of the unperturbed environment in which everyone plays α regardless of the observed signal.

Fact 1. *Let $\alpha \in \Delta(A)$ be a symmetric Nash equilibrium strategy of the underlying game $G=(A,\pi)$. Then the steady state $(S \equiv \{\alpha\}, 1_\alpha, \alpha)$ in which everyone plays α regardless of the observed signal is a Nash equilibrium in the unperturbed environment (G,k) for any $k \in \mathbb{N}$.*

A steady state is a (regularly) perfect equilibrium if it is the limit of Nash equilibria of (regular) perturbed environments when the frequency of the committed agents converges to zero. Formally, starting with standard definitions of convergence of a sequence of strategies and of a sequence of states, we have:

Definition 7. Fix environment (G,k). 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 $s_m(a)$. A sequence of states $(S_n,\sigma_n,\eta_n)_n$ converges to a state (S^*,σ^*,η^*) if for each pair of strategies $s,s' \in supp(\sigma^*)$, there exist sequences of strategies $(s_n)_n$ and $(s'_n)_n$ such that: (1) $s_n \to_{n\to\infty} s$ and $s'_n \to_{n\to\infty} s'$, (2) $\sigma_n(s_n) \to_{n\to\infty} \sigma^*(s)$ and $\sigma_n(s'_n) \to_{n\to\infty} \sigma^*(s')$, and (3) $(\eta_n)_{s_n}(s'_n) \to_{n\to\infty} (\eta^*)_s(s')$.

Definition 8. A steady state (S^*,σ^*,η^*) of the environment (G,k) is a *(regularly) perfect equilibrium* if there exist a (regular) distribution of commitments (S_C,λ) and converging sequences $(S_n,\sigma_n,\eta_n)_n \to_{n\to\infty} (S^*,\sigma^*,\eta^*)$ and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state (S_n,σ_n,η_n) is a Nash equilibrium of the perturbed environment $((G,k),(S_C,\lambda),\epsilon_n)$. In this case, we say that (S^*,σ^*,η^*) is a *perfect equilibrium with respect to distribution of commitments* (S_C,λ). If $\eta^* \equiv a$, we say that action $a \in A$ is a *perfect* equilibrium action.

By standard continuity arguments, any perfect equilibrium is a Nash equilibrium of the unperturbed environment. Next, observe that any symmetric perfect equilibrium α of the underlying game induces a corresponding perfect equilibrium of the environment in which all normal agents play α regardless of the observed signal. Moreover, if α is not totally mixed, then this steady state is a regular perfect equilibrium. Formally:

Proposition 1. *Let $\alpha \in A$ be a symmetric perfect equilibrium action of the underlying game $G=(A,\pi)$. Then the state $(S \equiv \{\alpha\}, 1_\alpha, \alpha)$ is a perfect equilibrium in the environment (G,k) for any $k \in \mathbb{N}$. Moreover, if the distribution α is not totally mixed, then $(S \equiv \{\alpha\}, 1_\alpha, \alpha)$ is a regular perfect equilibrium.*

An underlying game $G=((a,b),\pi)$ is a *(two-action) coordination game* if (a,a) and (b,b) are strict Nash equilibria. The next result shows that the totally mixed equilibrium of such a game does not induce a corresponding regular perfect equilibrium for any environment with[21] $k \geq 1$.

Proposition 2. *Let $G=(\{a,b\},\pi)$ be a coordination game. Let $\alpha \in \Delta(\{a,b\})$ be the mixed equilibrium action of G. Then the state $(S \equiv \{\alpha\}, 1_\alpha, \alpha)$ is not a regular perfect equilibrium in (G,k) for any $k \geq 1$.*

The intuition is that in the mixed equilibrium both actions earn the same expected payoff. The regularity of the set of commitment strategies implies there exists action a such that when an agent observes a sequence of only a:s, then the unique best reply is to play a because the partner is more likely to play a as well.

[21] The result can be extended to deal with the totally mixed equilibrium of a coordination game with more than two actions.

3.3 Strictly Perfect Equilibrium Action

In many of our results we focus on the stability of pure actions (i.e., cooperation and defection). The stability of a perfect equilibrium may crucially depend on a specific distribution of commitment strategies. The following definition of strict perfection of a pure outcome is much stronger, in the sense that it requires the pure outcome to be the limit behavior of Nash equilibria with respect to *all* commitment strategies.[22] Formally:

Definition 9. Action $a^* \in A$ is a *strictly perfect* equilibrium action in the environment $E = ((A, \pi), k)$, if for any distribution of commitment strategies (S_C, λ), there exist a steady state $(S^*, \sigma^*, \eta^* \equiv a^*)$ and converging sequences $(S_n, \sigma_n, \eta_n)_n \to_{n\to\infty} (S^*, \sigma^*, \eta^*)$ and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state (S_n, σ_n, η_n) is a Nash equilibrium of the perturbed environment $((G, k), (S_C, \lambda), \epsilon_n)$.

The following result shows that in an environment in which $k \geq 2$ and in which the game is a two-action coordination game, there is a unique strictly perfect equilibrium action, namely, the Pareto-efficient strict equilibrium action of the underlying game. This holds even if the Pareto-inefficient equilibrium is risk-dominant.[23],[24]

Proposition 3. *Let (G, k) be an environment where $G = ((a, b), \pi)$ is a coordination game and $k \geq 2$. The action a is a strictly perfect equilibrium action in the environment (G, k) if $\pi(a, a) > \pi(b, b)$, and it is not strictly perfect if*[25] *$\pi(a, a) < \pi(b, b)$.*

The essentially unique steady state that supports the Pareto-efficient action as a strictly perfect equilibrium action is similar to the steady state supporting cooperation in the defensive Prisoner's Dilemma in Theorem 2. It will be presented and discussed in Section 4.

The reason why the Pareto-dominated action (say, action a) is not a strictly perfect equilibrium action is the following. Consider a distribution of commitments that includes a commitment strategy that plays action b with high probability. Suppose all normal agents play action a with high probability. This means that if an agent observes a partner to always have played b then the partner is highly likely to be a commitment type who will continue to play b and hence the best response for a normal agent who receives a message of all b's is to play b. This implies that a deviator who always plays b induces all normal agents to play b, and thus she achieves a payoff of $\pi(b, b)$, which is strictly higher than the incumbents' average payoff (which is close to $\pi(a, a)$).

4 Analysis of the Prisoner's Dilemma

4.1 The Prisoner's Dilemma

In this section we focus on environments in which the underlying game is the Prisoner's Dilemma (denoted G_{PD}), which is described in Table 1 in the Introduction. The class of Prisoner's Dilemma games is fully described by 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 $-l$.

[22] We focus in the definition only on pure outcomes in order to simplify the notation (and because the stability of pure outcomes is the focus of many of our results). The definition is similar to the notion of strict perfection of Okada (1981).

[23] One can show that when $k = 1$ both pure actions are strictly perfect.

[24] The formal result deals with coordination games with two actions, but it can be extended to coordination games with more than two actions.

[25] In order to simplify the proof, we restrict attention to almost all commitment strategies (see Remark 7 in the proof). The proof can be extended to all commitment strategies, but as this would make the proof much lengthier, we omit it.

The fact that the Prisoner's Dilemma has two actions allows us to simplify the notation by setting $M = \{0, ..., k\}$, and interpreting $m \in M$ as the number of times that the partner defected in the sampled k observations.

Following Dixit (2003) we classify Prisoner's Dilemma games into two kinds: offensive and defensive.[26] In an *offensive* Prisoner's Dilemma 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 the following example.

Example 2. Consider a joint project of cowriting an academic paper in which each author can choose either to work hard (cooperate) or to shirk (defect). Assume that the joint paper is accepted (rejected) in a top journal for sure if both authors work hard (shirk), and that the paper is accepted with probability p if a single author works hard. Assume that publication in a top journal yields a benefit of 6 to each author, and working hard costs 5. If $p < 50\%$ the induced game is defensive (see the game G_D in Table 1 for the case $p = \frac{1}{3}$), while $p > 50\%$ induces an offensive game (see the game G_O in Table 1 for the case $p = 55\%$).

4.2 Stability of Defection and Anti-Folk Theorem

We begin by showing that defection is strictly perfect in any Prisoner's Dilemma game and for any k. Formally:

Proposition 4. *Let $E = (G_{PD}, k)$ be an environment. Defection is a strictly perfect equilibrium action.*

The intuition is straightforward. Consider any distribution of commitment strategies. Consider the stable state in which all the normal incumbents defect regardless of the observed signal. It is immediate that this strategy is the unique best reply to itself. This implies that if the share of committed agents is sufficiently small, then always defecting is also the unique best reply in the slightly perturbed environment.

Our first main result shows that defection is the *unique* regular perfect equilibrium in offensive games.

Theorem 1 (Anti-Folk Theorem)**.** *Let $E = (G_{PD}, k)$ be an environment, where G is an offensive Prisoner's Dilemma (i.e., $g > l$). If (S^*, σ^*, η^*) is a regular perfect equilibrium, then $(S^*, \sigma^*, \eta^*) = (\{d\}, 1_d, d)$.*

Sketch of Proof. The payoff of a strategy can be divided into two components: (1) a *direct* component: defecting yields additional g points if the partner cooperates and additional l points if the partner defects, and (2) an *indirect* component: the strategy's average probability of defection determines the distribution of signals 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 Prisoner's Dilemma 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). □

Discussion of Theorem 1 Our anti-folk theorem is surprising, as several papers in the existing literature present various mechanisms to support cooperation in any Prisoner's Dilemma game. Kandori (1992, Theorem 1) and Ellison (1994) show that cooperation can be supported by contagious equilibria even when an agent

[26] Takahashi (2010) calls offensive (defensive) Prisoner's Dilemmas submodular (supermodular).

does not observe any signal about her partner (i.e., $k = 0$). In these equilibria each agent starts the game by cooperating, but she starts defecting forever as soon as any partner has defected against her. Formally, such equilibria are impossible in our setup because the population is infinite rather than large and finite as in Kandori (1992) and Ellison (1994). Nevertheless, one can show that a small number of agents who follow regular commitment strategies destabilize all of the contagious equilibria also in large finite populations. Specifically (as pointed out by Ellison, 1994, p. 578), if we consider a large population in which at least one "crazy" agent defects with positive probability at all rounds regardless of the observed signal, then in such an environment no contagious equilibrium will be stable, because agents will assign high probability to the event that the contagion process has already begun, even after having experienced a long period during which no partner defected against them.

Sugden (1986) and Kandori (1992, Theorem 2) show that cooperation can be a stable equilibrium in a setup in which each player observes a binary signal about his partner, either a "good label" or a "bad label." All players start with a good label. This label becomes bad if a player defects against a "good" partner. The equilibrium strategy that supports full cooperation in this setup is to cooperate against good partners and defect against bad partners. Theorem 1 reveals that such a simple binary reputation cannot be maintained under an observation structure in which players observe an arbitrary number of past actions taken by their partners. The theorem shows this indirectly, because if it were possible to derive binary reputations from this information structure, then it should have been possible to prevent defection from being the unique perfect equilibrium action. Moreover, Theorem 4 in Section 5 shows that cooperation is not a perfect equilibrium in acute games when players observe action profiles. This suggests that the seemingly simple binary reputation mechanisms of Sugden (1986) and Kandori (1992) cannot be implemented under observation structures in which each agent observes the whole action profiles in the opponent's past interactions. We take this as a strong indication that it is difficult to maintain the required reputation system under plausible decentralized observation structures.

In Nowak and Sigmund's (1998) model of *image scoring* members of a community are randomly matched to play a one-sided helping game, in which an agent decides whether to help her partner by giving up g of her utility points in order to give $1 + g$ points to her partner. Each agent observes the last k actions taken by her partner. Nowak and Sigmund (1998) present simulation results suggesting that the image-scoring strategy, according to which an agent defects (refuses to help) iff the partner defected at least m times, might be an equilibrium. The one-sided helping game is closely related to the two-sided Prisoner's Dilemma with $g = l$ (in which, essentially, each agent decides giving up g of her utility points in order to yield $1 + g$ points to her partner). Our anti-folk theorem shows that it is impossible to extend image scoring to support cooperation in offensive Prisoner's Dilemmas (i.e., $g > l$). Theorem 2 shows that a novel variant of image scoring is essentially the unique mechanism to sustain cooperation in defensive Prisoner's Dilemmas, and Remark 5 in Section 4.3 discusses its extension to the case of $g = l$.

The mild restriction to a regular perfect equilibrium is necessary for the anti-folk theorem result to go through. The following example demonstrates the existence of a non-regular perfect equilibrium of an offensive PD, in which players cooperate with positive probability. This non-robust equilibrium is similar to the "belief-free" sequential equilibria that support cooperation in offensive Prisoner's Dilemma games in Takahashi (2010) (see also Deb, 2012), which have the property that players are always indifferent between their actions, but they choose different mixed actions depending on the signal they obtain about the partner. The following example also illustrates why all of Takahashi's (2010) equilibria crucially depend on the absence of commitment strategies, and why the presence of any arbitrarily small group of regular commitment strategies imply the anti-folk theorem. Elsewhere, Heller (2015a) adapts a general non-robustness argument related to the example

above to the standard setup of repeated games played by the same two players, and shows that also in this standard setup, none of the belief-free equilibria are robust against small perturbations in the behavior of potential opponents.

Example 3 (Non-regular Perfect Equilibrium with Partial Cooperation)**.** Consider the environment $(G_O, 1)$ where G_O is an offensive Prisoner's Dilemma game with $g = 2.3$, $l = 1.7$ (see Table 1), and each agent observes a single action sampled from the partner's behavior. Let s^* be the strategy that defects with probability 10% after observing cooperation (i.e., $m = 0$) and defects with probability 81.7% (numerical values in this example are rounded to 0.1%) after observing a defection (i.e., $m = 1$). Let q^* denote the average probability of defection in a homogeneous population of agents who follow strategy s^*. The value of q^* is calculated as follows:

$$q^* = (1 - q^*) \cdot 10\% + q^* \cdot 81.7\% \;\Rightarrow\; q^* = 35.3\%. \tag{8}$$

Eq. (8) holds because an agent defects in either of the following exhaustive cases: (1) she observes cooperation (which happens with a probability of $1 - q^*$) and then she defects with probability 10%, or (2) she observes defection (which happens with a probability of q^*) and then she defects with probability 81.7%. This implies that $\eta^* \equiv 35.3\%$ is the unique consistent outcome of a homogeneous population in which all agents follow s^*. Next, observe that that an agent who follows strategy s^* defects with probability

$$p(q) = q \cdot 81.7\% + (1 - q) \cdot 10\%$$

when being matched with a partner who defects with an average probability of q. This implies that the payoff of a deviator (Alice) who defects with an average probability of q is:

$$\pi_q\left(\left(\{s^*\}, 1_{s^*}, \eta^* \equiv 35.3\%\right)\right) = q \cdot (1 - p(q)) \cdot (1 + g) + (1 - q) \cdot p(q) \cdot (-l) + (1 - q) \cdot (1 - p(q)) \cdot 1.$$

This is because with a probability of $q \cdot (1 - p(q))$ only Alice defects, with a probability of $(1 - q) \cdot p(q)$ only Alice cooperates, and with a probability of $(1 - q) \cdot (1 - p(q))$ both players cooperate. By calculating the FOC one can show that $q = q^* = 35.3\%$ is the probability of defection that uniquely maximizes the payoff of a deviator. This implies that $(\{s^*\}, 1_{s^*}, 35.3\%)$ is a Nash equilibrium of the (non-regular) perturbed environments $(G, k, \{s^*\}, 1_{s^*}, \epsilon)$ for any $\epsilon \in (0, q)$, which implies that $(\{s^*\}, 1_{s^*}, 35.3\%)$ is a (non-regular) perfect equilibrium. The above perfect equilibrium relies on a very particular set of commitment strategies in which all committed agents happen to play the same strategy as the normal agents. This cannot hold in a regular set of commitment strategies, in which different commitment strategies defect with different average probabilities. Given this regularity, it must be the case that the conditional probability that the partner is going to defect is higher after he observes a defection ($m = 1$), than after he observes a cooperation ($m = 0$). This implies that a deviator (Alice) who defects with a probability of 35.3% regardless of the signal will strictly outperform the incumbents. This is because the incumbents behave the same against Alice (as she has the same average probability of defection as the incumbents), while Alice defects with higher probability against partners who are more likely to cooperate (i.e., after she observes $m = 0$), which implies that due to the offensiveness of the game (i.e., $g > l$), Alice achieves a strictly higher payoff than the incumbents.

4.3 Stability of Cooperation in Defensive Prisoner's Dilemmas

Our second main result shows that if players observe at least two actions, then cooperation is strictly perfect in any defensive Prisoner's Dilemma. Moreover, it shows that there is essentially a unique combination of strategies

that supports (full) cooperation in the Prisoner's Dilemma game, according to which: (a) all agents cooperate when observing no defections, (b) all agents defect when observing at least 2 defections, (3) sometimes (but not always) agents defect when observing a single defection. The average defection probability when an agent observes a single defection depends on the strategy commitments, and it is in the interval $\left[\frac{g}{l+1}\cdot\frac{1}{k},\frac{l}{l+1}\cdot\frac{1}{k}\right]$. Formally:

Theorem 2. *Let $E=(G_{PD},k)$ be an environment with observations of actions, where G_{PD} is a defensive Prisoner's Dilemma ($g<l$), and $k\geq 2$.*

1. *If $(S^*,\sigma^*,\eta^*\equiv c)$ is a perfect equilibrium then: (a) for each $s\in S^*$, $s_0(c)=1$ and $s_m(d)=1$ for each $m\geq 2$; and (b) there exist $s,s'\in S^*$ such that $s_1(d)<1$ and $s'_1(d)>0$.*

2. *Cooperation is a strictly perfect equilibrium action.*

Sketch of Proof. Suppose that $(S^*,\sigma^*,\eta^*\equiv c)$ is a perfect equilibrium. The fact that the equilibrium induces full cooperation, in the limit when the mass of commitment strategies converges to zero, implies that all normal agents must cooperate when they observe no defections, i.e., $s_0(c)=1$ for each $s\in S^*$.

Next we show that there is a normal strategy that induces the agent to defect with positive probability when observing a single defection, i.e., $s_1(d)>0$ for some $s\in S^*$. Assume to the contrary that $s_1(c)=1$ for each $s\in S^*$. If an agent (Alice) deviates and defects with small probability $\epsilon<<1$ when observing no defections, then she outperforms the incumbents. On the one hand, the fact that she rarely defects when observing $m=0$ gives her a direct gain of at least $\epsilon\cdot g$. On the other hand, the probability that a partner observes her defecting twice or more is $O\left(\epsilon^2\right)$, thus her indirect loss from these additional ϵ defections is at most $O\left(\epsilon^2\right)\cdot(1+l)$, and thus for a sufficiently small $\epsilon>0$ Alice strictly outperforms the incumbents.

The fact that $s_1(d)>0$ for some $s\in S^*$ implies that defection is a best reply conditional on observing $m=1$. The direct gain from defecting is strictly increasing in the probability that the partner defects (because the game is defensive), while the indirect influence of defection on the behavior of future partners is independent of the partner's play. This implies that defection must be the unique best reply when an agent observes $m\geq 2$ defections, since such an observation implies a higher probability that the partner is going to defect relative to the observation of a single defection. This establishes $s_m(d)=1$ for all $m\geq 2$ and all $s\in S^*$.

In order to show that $s_1(d)<1$ for $s\in S^*$, assume to the contrary that $s_1(d)=1$ for each $s\in S^*$. One can show that this implies that the unique consistent behavior is $\eta^*\equiv d$. The intuition is that the defection of all incumbents when observing a single defection implies that defections (which always occur due to the presence of commitment types) are "contagious": any defection of an agent induces at least one partner to defect against the agent, and thus the unique consistent behavior must be such that agents always defect. This completes the sketch of the proof of part (1).

Let s^1 and s^2 be the strategies that defect iff $m\geq 1$ and $m\geq 2$, respectively. Consider the state $\left(\left\{s^1,s^2\right\},(q^*,1-q^*),\eta^*\equiv c\right)$. The direct gain from defecting (relative to cooperating) when observing $m=1$ is

$$\Pr(m=1)\cdot\left(\left(l\cdot\Pr(d|m=1)\right)+g\cdot\Pr(c|m=1)\right),$$

where $\Pr(d|m=1)$ ($\Pr(c|m=1)$) is the probability that a random partner is going to defect (cooperate) conditional on the agent observing $m=1$, and $\Pr(m=1)$ is the average probability of observing the signal $m=1$. The indirect loss from defection, relative to cooperation, conditional on the agent observing a single defection, is

$$q^*\cdot(k\cdot\Pr(m=1))\cdot(l+1)+O\left(\left(\Pr(m=1)\right)^2\right);$$

this is because a random partner defects with an average probability of q if he observes a single defection (which occurs with probability $k \cdot \Pr(m=1)$ when the partner has k i.i.d. observations, each of which has a probability of $\Pr(m=1)$ of being a defection), and each induces a loss of $l+1$ to the agent (who obtains $-l$ instead of 1). The fact that some normal agents cooperate and others defect when observing a single defection, implies that both actions are best replies conditional on observing $m=1$. This implies that the indirect loss from defecting is exactly equal to the direct gain (up to $O\left(\left(\Pr(m=1)\right)^2\right)$), which is:

$$\Pr(m=1)\cdot((l\cdot\Pr(d|m=1))+g\cdot\Pr(c|m=1))=q^*\cdot(k\cdot\Pr(m=1))\cdot(l+1)$$

$$\Rightarrow q^*=\frac{(l\cdot\Pr(d|m=1))+g\cdot\Pr(c|m=1)}{k\cdot(l+1)}. \tag{9}$$

The probability $\Pr(d|m=1)$ depends on the distribution of commitments. Yet, one can show that for any distribution of commitment strategies (S_C,λ), there is a unique value of $q^*\in(0,1)$ that solves Eq. (9) and that, given this q^*, both s^1 and s^2 (and only these strategies) are best replies. This means that $\left(\left\{s^1,s^2\right\},(q^*,1-q^*),\eta^*\equiv c\right)$ is a perfect equilibrium. □

Discussion of Theorem 2 We comment on a few issues related to Theorem 2.

1. Each distribution of commitment strategies induces a unique frequency q^* of s^1-agents that yields a perfect equilibrium. One may wonder whether a population starting from a different share $q_0\neq q^*$ of s^1-agents is likely to converge to the "correct" frequency q^*. It is possible to show that the answer is affirmative. Specifically, given any initial low frequency $q_0\in(0,q^*)$, the s^1-agents achieve a higher payoff than the s^2-agents and, given any initial high frequency $q_0\in\left(q^*,\frac{1}{k}\right)$, the s^1-agents achieve a lower payoff than the s^2-agents. Thus, under any smooth monotonic dynamic process in which a more successful strategy gradually becomes more frequent, the share of s^1-agents will shift from any initial value in the interval $q_0\in\left(0,\frac{1}{k}\right)$ to the exact value of q^* that induces a perfect equilibrium.

2. As discussed in the formal proof in Appendix B, some distributions of commitment strategies may induce a slightly different perfect equilibrium, in which the population is homogeneous, and each agents in the population defects with probability q^* (μ) when observing a single defection (contrary to the heterogeneous deterministic behavior described above).

3. In Appendix B we show that the stability of cooperation is robust to small group of agents (with a positive small mass) who jointly deviate (à la Maynard Smith and Price's (1973) notion of evolutionary stability).

4. Our results can be extended to a setup in which the number of observed actions is random. Specifically, consider a *random environment* (G_{PD},p), where $p\in\Delta(N)$ is a distribution with a finite support, and each agent privately observes k actions of the partner with probability $p(k)$. Specifically, Theorem 2 (and, similarly, Theorems 3–5) will hold for any random environment in which the probability of observing at least two interactions is sufficiently high. The perfect equilibrium has to be adapted as follows. As in the main model, all normal agents cooperate (defect) when observing no (at least two) defections. In addition, there will be a value $\bar{k}\in supp(p)$ and a probability $q\in[0,1]$ (which depend on the distribution of commitment strategies), such that all normal agents cooperate (defect) when observing a single defection out of $k>\bar{k}$ $(k<\bar{k})$, and a fraction of q of the normal agents defect when observing a single defection out of $\bar{k}$ observations.

5. The threshold case between defensiveness and offensiveness: $g = l$. Such a Prisoner's Dilemma game can be interpreted as a game in which each of the players simultaneously decides whether to sacrifice a personal payoff of g in order to induce a gain of $1 + g$ to her partner. One can show that cooperation is also strictly perfect in this setup, and it is supported by the same kind of perfect equilibrium as described above. However, in this case: (I) the uniqueness result (part (1) of Theorem 2) is no longer true, as other kinds of strategies may also support full cooperation, and (II) cooperation does not satisfy the refinement of evolutionary stability (Appendix B). One can adapt the proof of Theorem 1 to show that defection is the unique perfect evolutionarily stable outcome when $g = l$.

The following example demonstrates the existence of a perfect equilibrium that supports cooperation when the unique commitment strategy is to play each action uniformly.

Example 4 (Example 1 revisited: Illustration of the perfect equilibrium that supports cooperation)**.** Consider the perturbed environment $(G_D, 2, \{s^u \equiv 0.5\}, 1_{s^u}, \epsilon)$, where G_D is the defensive Prisoner's Dilemma game with the parameters $g = 1$ and $l = 3$ (as presented in Table 1 in the Introduction). Consider the stable state $\left(\left\{s^1, s^2\right\}, \left(\frac{1}{6}, \frac{5}{6}\right), \eta^*\right)$, where η^* is defined as in (6) in Example **1** above. A straightforward calculation shows that the average probability in which a normal agent observes $m = 1$ when being matched with a random partner is

$$\Pr(m = 1) = \epsilon \cdot 0.5 + 3.5 \cdot \epsilon \cdot \frac{1}{6} + 0.5 \cdot \epsilon \cdot \frac{5}{6} + O\left(\epsilon^2\right) = 1.5 \cdot \epsilon + O\left(\epsilon^2\right).$$

The probability that the partner is a committed agent conditional on observing a single defection is:

$$\Pr(s^u | m = 1) = \frac{\epsilon \cdot 0.5}{1.5 \cdot \epsilon} = \frac{1}{3} \;\Rightarrow\; \Pr(d | m = 1) = \frac{1}{3} \cdot 0.5 = \frac{1}{6},$$

which yields the conditional probability that the partner of a normal agent will defect. Next we calculate the direct gain from defecting conditional on the agent observing a single defection ($m = 1$):

$$\Pr(m = 1) \cdot ((l \cdot \Pr(d | m = 1)) + g \cdot \Pr(c | m = 1)) = 1.5 \cdot \epsilon \cdot \left(3 \cdot \frac{1}{6} + 1 \cdot \frac{5}{6}\right) + O\left(\epsilon^2\right) = 2 \cdot \epsilon + O\left(\epsilon^2\right).$$

The indirect loss from defecting conditional on the agent observing a single defection is:

$$q \cdot (k \cdot \Pr(m = 1)) \cdot (l + 1) + O\left(\epsilon^2\right) = q \cdot 2 \cdot 1.5 \cdot \epsilon \cdot (3 + 1) = 12 \cdot q \cdot \epsilon + O\left(\epsilon^2\right).$$

When taking $q = \frac{1}{6}$ the indirect loss from defecting is exactly equal to the direct gain (up to $O\left(\epsilon^2\right)$).

4.4 Stability of Cooperation when Observing a Single Action

Given a distribution of commitments (S_C, λ), we define $\beta_{(S_C,\lambda)} \in (0, 1)$ as follows:

$$\beta_{(S_C,\lambda)} = \frac{\mathbf{E}_\lambda\left(\left(s_0(d)\right)^2\right)}{\mathbf{E}_\lambda\left(\left(s_0(d)\right)\right)} = \frac{\sum_{s \in S_C} \lambda(s) \cdot \left(s_0(d)\right)^2}{\sum_{s \in S_C} \lambda(s) \cdot s_0(d)}. \tag{10}$$

The value of $\beta_{(S_C,\lambda)}$ is the ratio between the mean of the square of the probability of defection of a random committed agent who observes $m = 0$, and the mean of the same probability without squaring it. In particular, when the set of commitments is a singleton, $\beta_{(S_C,\lambda)}$ is equal to the probability that a committed agent defects when she observes $m = 0$ (i.e., $\beta_{(S_C,\lambda)} = s_0(d)$).

The following result shows that if the game is defensive and agents observe a single action, then full cooperation is a perfect equilibrium action with respect to distribution of commitments (S_C, λ) iff $g \leq \beta_{(S_C,\lambda)}$. In particular, cooperation is a strictly perfect equilibrium iff $g < 1$.

Proposition 5. *Let $E = (G_{PD}, 1)$ be an environment, where G_{PD} is a defensive Prisoner's Dilemma ($g < l$). Let (S_C, λ) be a distribution of commitments. There exists a perfect equilibrium $(S^*, \sigma^*, \eta^* \equiv c)$ with respect to (S_C, λ) iff $g \leq \beta_{(S_C,\lambda)}$.*

Sketch of Proof. Similar arguments to those presented in part (1) of Theorem 2 imply that any distribution of commitment strategies induces a unique average probability q by which normal agents defect when observing $m = 1$, in any cooperative perfect equilibrium. This implies that a deviator who always defects gets a payoff of $1 + g$ in a fraction $1 - q$ of the interactions. One can show that such a deviator outperforms the incumbents iff[27] $g > \beta_{(S_C,\lambda)}$. □

Corollary 1. *Let $E = (G_{PD}, 1)$ be an environment, where G_{PD} is a defensive Prisoner's Dilemma ($g < l$). Cooperation is a strictly perfect equilibrium action iff $g \leq 1$.*

The intuition for the difference compared to the case of $k \geq 2$ is that the higher g is, the more severely defection has to be punished. However, since $k = 1$, defection can only be deterred by increasing the probability of defecting upon observing $m = 1$, and this deterrent is not enough if g is sufficiently high.

5 General Observation Structures

In this section we extend our analysis to general observation structures in which the signal about the partner may also depend on the behavior of other opponents against the partner.

5.1 Definitions

An *observation structure* is a tuple $\Theta = (k, B, o)$, where $k \in \mathbb{N}$ is the number of observed interactions, $B = \{b_1, .., b_{|B|}\}$ is a finite set of *observations* that can be made in 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). Note that observing actions (which was analyzed in the previous section) is equivalent to having $B = A$ and $o(a, a') = a$.

In the results of this section we focus on three observation structures:

1. *Observation of action profiles:* $B = A^2$ and $o(a, a') = (a, a')$. In this observation structure, each agent observes, in each sampled interaction of her partner, both the action played by her partner and the action played by her partner's opponent.

2. *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 $(a, a') \neq (c, c)$. Such an observation structure (which we have not seen in the existing literature) seems like a plausible way to capture non-verifiable feedback about the partner's behavior. The agent can observe in each sampled past interaction of the partner if both partners were "happy" (i.e., mutual cooperation), or if the partners complain about each other (i.e.,

[27] In environments with $k \geq 2$, a deviator who always defects gets a payoff of zero, regardless of the value of q (because all agents observe $m = k$ when being matched with such a deviator).

there was a conflict, at least one of the players defected, and it is too costly for an outside observer to verify who actually defected).

3. Observation of actions against cooperation: $B = \{CC, DC, *D\}$ and $o(c,c) = CC$, $o(d,c) = DC$, and $o(c,d) = o(d,d) = *D$. That is, each agent (Alice) observes a ternary signal about each sampled interaction of her partner (Bob): either both players cooperated, or Bob unilaterally defected, or Bob's partner defected (and in this latter case Alice cannot observe Bob's action). We analyze this observation structure because it turns out to be an "optimal" observation structure that allows cooperation to be supported as a perfect equilibrium action in any Prisoner's Dilemma.

In each of these cases, we let the mapping o be implied by the context, and identify the observation structure Θ with the number of observed interactions k.

In what follows we present the definitions of the main model (Section 2–3) that have to be changed to deal with the general observation structure. Before playing the game, each player independently samples k independent interactions of her partner. Let M denote the set of feasible signals:

$$M = \left\{ m \in \mathbb{N}^{|B|} \middle| \sum_i m_i = k \right\},$$

where m_i is interpreted as the number of times that observation b_i has been observed in the sample. When the underlying game is the Prisoner's Dilemma and agents observe conflicts, we simplify the notation by identifying $M = \{1, ..., k\}$, and interpreting $m \in \{1, ..., k\}$ as the number of observed conflicts.

The definition of a strategy remains the same (with the new set of feasible messages M). In particular, recall that for each strategy $s \in \mathcal{S}$ and distribution of signals $\nu \in \Delta(M)$, we define $s_\nu \in \Delta(A)$ to be the distribution of actions played by an agent who follows strategy s and observed a signal sampled from ν.

A perturbed environment is a tuple $E_\epsilon = ((G, \Theta), (S_C, \lambda), \epsilon)$, where Θ is an observation structure, and all other components are the same as in Definition 2. The consistency requirement in the definition of a steady state has to be adapted to the general observation structure as follows.

Definition 10 (Adaptation of Def. 4)**.** A *steady state* (or *state*) of a perturbed environment $((G, k), (S_C, \lambda), \epsilon)$ is a triple (S, σ, η), where $S \subseteq \mathcal{S}$ is a finite set of strategies, $\sigma \in \Delta(S)$ is a distribution, and $\eta : (S \cup S_C) \times (S \cup S_C) \to \triangle(A)$ is a mapping (called, *consistent behavior*) that assigns to each pair of strategies $s, s' \in S \cup S_C$ a mixed action $\eta_s(s') \in \Delta(A)$ and satisfies the consistency condition below. Let $\psi_s \in \triangle(A \times A)$ be the (possibly correlated) mixed action profile that is played when an agent with strategy $s \in S \cup S_C$ is matched with a random partner (given σ and η). 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 her partner,

$$\psi_s(a, a') = \sum_{s' \in S \cup S_C} ((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')) \cdot \eta_s(s')(a) \cdot \eta_{s'}(s)(a').$$

Let $\nu_s \in \Delta(M)$ be the distribution of signals induced by an agent who follows strategy s (given σ and η).

$$\forall (m_1, ..., m_{|B|}) \in M, \quad \nu_s(m_1, ..., m_{|B|}) = \left(k! \cdot \prod_{1 \le i \le |B|} \frac{\left(\sum_{(a,a') \in A \times A} (\psi_s(a, a') \cdot (o(a, a')(b_i))) \right)^{m_i}}{m_i!} \right). \quad (11)$$

The *consistency requirement* that the mapping η has to satisfy is

$$\forall a \in A,\ s, s' \in S \cup S_C, \quad \eta_s(s')(a) = s_{\nu_{s'}}(a). \tag{12}$$

As in the main model, the interpretation of the consistency requirement (12) is as follows. When Alice (who follows s) is being matched with Bob (who follows s'), she observes each signal m with a probability of $\nu_{s'}(m)$, and conditional on observing signal m, she plays each action a with a probability of $s_m(a)$. Furthermore, as in the main model, a standard fixed-point argument shows that any distribution of strategies admits a consistent behavior (which is not necessarily unique). The definitions of the long-run payoff of an incumbent who follows a stationary strategy $s \in S \cup S_C$, $\pi_s(S, \sigma, \eta)$ and the definition of the average payoff of the normal agents, $\pi(S, \sigma, \eta)$ remains unchanged.

We now adapt the definition of the payoff of an agent (Alice) who deviates and plays a new (non-incumbent) strategy $\hat{s} \in \mathcal{S} \setminus (S \cup S_C)$. Unlike in the basic model, in this extension there might be multiple consistent outcomes following Alice's deviation, as demonstrated in Example 5.

Example 5. Consider an unperturbed environment $(G_{PD}, 3)$ with observation of $k = 3$ action profiles. Consider a homogeneous incumbent population in which all agents follow the following strategy: $s^*(m) = d$ if m includes at least 2 interactions with (d, d), and $s^*(m) = c$ otherwise. Consider the state $(\{s^*\}, 1_{s^*}, \eta^* = c)$ in which everyone cooperates. Consider a deviator (Alice) who follows the strategy of always defecting. Then there exist three consistent post-deviation steady states (in all of which the incumbents continue to cooperate among themselves): (1) all the incumbents defect against Alice, (2) all the incumbents cooperate against Alice, and (3) all the incumbents defect against Alice with a probability of 50%.

We define the payoff of a deviator as the highest payoff she may get in any post-deviation steady state.

Definition 11 (Adaptation of Def. 5)**.** Given steady state (S, σ, η) in a perturbed environment $(G, \Theta, S_C, \lambda, \epsilon)$ and strategy $\hat{s} \in \mathcal{S} \setminus (S \cup S_C)$, we say that $(S \cup \{\hat{s}\}, \hat{\sigma}, \hat{\eta})$ is a *post-deviation steady-state* if it satisfies:

1. $\hat{\sigma}(s) = \sigma(s)$ for each $s \in S$ and $\hat{\sigma}(\hat{s}) = 0$.
2. $\hat{\eta}_s(s') = \eta_s(s')$ for each $s, s' \in S \cup S_C$.
3. $\forall a \in A,\ s' \in S \cup S_C \quad \hat{\eta}_{\hat{s}}(s')(a) = \hat{s}_{\nu_{s'}}(a)$ (where $\nu_{s'}$ is defined as in (11) above).
4. Let $\hat{\psi}_{\hat{s}} \in \triangle(A \times A)$ be the (possibly correlated) mixed action profile that is played when the deviator is matched with a random partner (given $\hat{\sigma}$ and $\hat{\eta}$):

$$\hat{\psi}_{\hat{s}}(a, a') = \sum_{s' \in S \cup S_C} ((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')) \cdot \hat{\eta}_{\hat{s}}(s')(a) \cdot \hat{\eta}_{s'}(\hat{s})(a').$$

Let $\hat{\nu}_{\hat{s}} \in \Delta(M)$ be the distribution of signals induced by the deviator (given $\hat{\sigma}$ and $\hat{\eta}$):

$$\forall (m_1, ..., m_{|B|}) \in M, \quad \hat{\nu}_{\hat{s}}(m_1, ..., m_{|B|}) = \left(k! \cdot \prod_{1 \le i \le |B|} \frac{\sum_{(a,a') \in A \times A} (\psi_{\hat{s}}(a, a') \cdot (o(a, a')(b)))}{m_i!} \right).$$

Then the final consistency requirement is: $\forall a \in A,\ s' \in S \cup S_C \quad \hat{\eta}_{s'}(\hat{s})(a) = s_{\hat{\nu}_{\hat{s}}}(a)$.

Let $\pi_{\hat{s}}(S \cup \{\hat{s}\}, \hat{\sigma}, \hat{\eta})$ be the payoff of the deviator who follows $\hat{s}$ in the post-deviation state $(S \cup \{\hat{s}\}, \hat{\sigma}, \hat{\eta})$. Let $PESS(((S, \sigma, \eta)), \hat{s})$ be the set of all post-entry deviation states (which is non-empty and compact by

standard arguments). Let $\pi_{\hat{s}}(S,\sigma,\eta)$ be the maximal payoff for a deviator in a post-deviation steady state:

$$\pi_{\hat{s}}(S,\sigma,\eta) :=_{(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta})\in PESS(((S,\sigma,\eta)),\hat{s})} \max \pi_{\hat{s}}(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta}). \tag{13}$$

Remark 3. Our results remain the same if one replaces the maximum function in (13) with a minimum function.

5.2 Acute and Mild Prisoner's Dilemma

In this subsection we present a novel classification of Prisoner's Dilemma games that plays an important role in the results of this section. Recall that the parameter g of a Prisoner's Dilemma game may take any value in the interval $[0, l+1]$ (if $g > l+1$, then mutual cooperation is no longer the efficient outcome that maximizes the sum of payoffs). We say that a Prisoner's Dilemma game 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 (i.e., if $g < \frac{l+1}{2}$). 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 games and *acutely tempting* in acute games. In order for an agent not to be tempted to defect against a cooperating partner in an acute (one-shot) Prisoner's Dilemma, 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 games requires this probability of being punished to be more than 50%, while a probability of below 50% is enough for mild games.

Example 6. Table 3 demonstrates the payoffs of specific acute (G_A) and mild (G_M) Prisoner's Dilemma games. In both examples $g = l$, i.e., the Prisoner's Dilemma game is "linear." This means that it can be described as a "helping game" in which agents have to decide simultaneously whether to give up a payoff of g in order to create a benefit of $1+g$ for the partner. In the acute game (G_A) on the left, $g = 3$ and the loss of a helping player amounts to more than half of of the benefit to the partner who receives the help ($\frac{3}{3+1} = \frac{3}{4} > \frac{1}{2}$), while in the mild game (G_M) on the right, $g = 0.2$ and the loss of the helping player is less than half of the benefit to the partner who receives the help ($\frac{0.2}{0.2+1} = \frac{1}{6} < \frac{1}{2}$).

Table 3: Matrix Payoffs of Acute and Mild Prisoner's Dilemma Games

	c	d
c	1 1	$-l$ $1+g$
d	$1+g$ $-l$	0 0

General Prisoner's Dilemma
G_{PD}: $g, l > 0$, $g < l+1$

	c	d
c	1 1	-3 4
d	4 -3	0 0

Ex. 3: Acute Prisoner's Dilemma
G_A: $g = l = 3 > \frac{l+1}{2} = 2$

	c	d
c	1 1	-0.2 1.2
d	1.2 -0.2	0 0

Ex. 4: Mild Prisoner's Dilemma
G_M: $g = l = 0.2 < \frac{l+1}{2} = 0.6$

5.3 Analysis of the Stability of Cooperation

We first note that Proposition 4 is valid also in this extended setup with minor adaptations to the proof. Thus, always defecting is a strictly perfect equilibrium regardless of the observation structure. Next we analyze the stability of cooperation in each of the three interesting observation structures.

The following two results show that under either **observation of conflicts** or **observation of action profiles**, cooperation is a perfect equilibrium iff the Prisoner's Dilemma is mild ($g < 0.5 \cdot (l+1)$) . Moreover, in mild Prisoner's Dilemma games there is essentially a unique strategy distribution that supports cooperation (which is analogous to the essentially unique strategy distribution in Theorem 2). Formally:

Theorem 3. *Let $E = (G, p)$ be an environment with observation of conflicts, where G is a PD and $k \geq 2$.*

1. *If G is a mild PD ($g < \frac{l+1}{2}$), then:*
   (a) *If $(S^*, \sigma^*, \eta^* \equiv c)$ is a perfect equilibrium then: (I) for each $s \in S^*$, $s_0(c) = 1$ and $s_m(d) = 1$ for each $m \geq 2$; and (II) there exist $s, s' \in S^*$ such that $s_1(d) < 1$ and $s'_1(d) > 0$.*
   (b) *Cooperation is a strictly perfect equilibrium action.*
2. *If G is an acute PD ($g > \frac{l+1}{2}$), then cooperation is not a perfect equilibrium action.*

Sketch of proof. The argument for part (1a) is analogous to Theorem 2. In what follows we sketch the proofs of part 1(b) and part 2. Fix a distribution of commitments, and a commitment level $\epsilon \in (0,1)$. Let m denote the number of observed conflicts and define s^1 and s^2 as before, but with the new meaning of m. Consider the following candidate for a perfect equilibrium $\left(\left\{s^1, s^2\right\}, (q, 1-q), c\right)$. Here, the probability q will be determined such that both actions are best replies when observing a single conflict. That is, the direct benefit from defecting when observing $m = 1$ (the LHS of the equation below) must balance the indirect loss due to inducing future partners who observe these conflicts to defect (the RHS, neglecting terms of $O(\epsilon)$). The RHS is calculated by noting that defection induces an additional conflict only if the current partner has cooperated and that, on expectation, each such additional conflict is observed by k future partners, each of whom defects with an average probability of q). Recall that $\Pr(d|m=1)$ ($\Pr(c|m=1)$) is the probability that a random partner is going to defect (cooperate) conditional on the agent observing $m = 1$.

$$\Pr(m=1) \cdot ((l \cdot \Pr(d|m=1)) + g \cdot \Pr(c|m=1)) = \Pr(m=1) \cdot k \cdot q \cdot \Pr(c|m=1) \cdot (l+1)$$
$$\Leftrightarrow q \cdot k = \frac{(l \cdot \Pr(d|m=1)) + g \cdot \Pr(c|m=1)}{\Pr(c|m=1) \cdot (l+1)}. \tag{14}$$

One can see that the RHS is increasing in $\Pr(d|m=1)$. The minimal bound on the value of q is obtained when $\Pr(d|m=1) = 0$. In this case $q \cdot k = \frac{g}{l+1}$.

Suppose the game is acute. In this case $q \cdot k > 0.5$. Suppose that the average probability of defection in the population is $\Pr(d)$. Since there is full cooperation in the limit we have $\Pr(d) = O(\epsilon)$. This implies that a fraction $2 \cdot \Pr(d) + O(\epsilon^2)$ of the population is involved in conflicts. This in turn induces the defection of a fraction $2 \cdot \Pr(d) \cdot k \cdot q + O(\epsilon^2)$ of the normal agents (because a normal agent defects with probability q upon observing at least one conflict in the k sampled interactions). Since the normal agents constitute a fraction $1 - O(\epsilon)$ of the population we must have $\Pr(d) = 2 \cdot \Pr(d) \cdot k \cdot q + O(\epsilon^2)$. However, in an acute game, $2 \cdot k \cdot q > 1$ leads to the contradiction that $\Pr(d) < \Pr(d)$. Thus, if $2 \cdot k \cdot q > 1$, then defections are contagious, and so there is no steady state in which only a fraction $O(\epsilon)$ of the population defects.

Suppose the game is mild. One can show that $\Pr(d|m=1)$ is decreasing in q, and that it converges to zero when $k \cdot q \nearrow 0.5$. (The reason is that when $k \cdot q$ is close to 0.5 each defection by a committed agent induces many defections by normal agents and, conditional on observing $m = 1$, the partner is likely to be normal and to cooperate when being matched with a normal agent.) It follows that the RHS of Eq. (14), is decreasing in q and approaches the value $\frac{g}{l+1}$ when $k \cdot q \nearrow 0.5$. Since the game is mild, $\frac{g}{l+1} < 0.5$. Hence there is some $q \cdot k < 0.5$ that solves Eq. (14), and in which the normal agents defect with a low probability of $(O(\epsilon))$. □

Theorem 4. *Let $E = (G_{PD}, k)$ be an environment with observation of action profiles and $k \geq 2$. Cooperation is a perfect equilibrium action iff the underlying game G_{PD} is mild (i.e., $g \leq \frac{l+1}{2}$).*

Sketch of proof. Using arguments that are familiar from above one can show that in any perfect equilibrium that supports cooperation, normal agents have to defect with an average probability of $q \in (0,1)$ when observing a single unilateral defection (and $k-1$ mutual cooperations), and defect with a smaller probability when observing a single mutual defection (since this is necessary in order for a normal agent to have better incentives to cooperate against a partner who is more likely to cooperate). The value of q is determined by Eq. (14) above, implying that both actions are best replies conditional on observing the partner to be the sole defector once, and to be involved in mutual cooperation in the remaining $k-1$ observed action profiles. Let ϵ be the share of committed agents, and let φ be the average probability that a committed agent unilaterally defects. In order to simplify the sketch of the proof, we will focus on the case in which the committed agents defect with a small probability when observing the partner to have been involved only in mutual cooperations, which implies, in particular, that $\varphi << 1$ (the formal proof in the Appendix does not make this simplifying assumption). The unilateral defections of the committed agents induce a fraction of $\epsilon \cdot \varphi \cdot k \cdot q + O\left(\epsilon^2\right) + O\left(\varphi^2\right)$ of the normal agents to defect when being matched against committed agents (because a normal agent defects with probability q upon observing a single unilateral defection in the k sampled interactions). These unilateral defections of normal agents against committed agents induce a further $(\epsilon \cdot \varphi \cdot k \cdot q) \cdot k \cdot q + O\left(\epsilon^2\right)$ defections of normal agents against other normal agents. Repeating this argument we come to the conclusion that the average probability of a normal agent being the sole defector is (neglecting terms of $O\left(\epsilon^2\right)$ and $O\left(\varphi^2\right)$):

$$\epsilon \cdot \varphi \cdot k \cdot q \cdot \left(1 + k \cdot q + (k \cdot q)^2 + ...\right) = \epsilon \cdot \varphi \cdot \frac{k \cdot q}{1 - k \cdot q.}$$

As discussed above, in acute games, the value of $k \cdot q$ must be larger than 0.5, which implies that $\frac{k \cdot q}{1-k \cdot q} > 1$. This implies that conditional on observing the partner to be the sole defector once, the posterior probability that the partner is normal is:

$$\frac{\epsilon \cdot \varphi \cdot \frac{k \cdot q}{1-k \cdot q}}{\epsilon \cdot \varphi + \epsilon \cdot \varphi \cdot \frac{k \cdot q}{1-k \cdot q}} = \frac{\frac{k \cdot q}{1-k \cdot q}}{1 + \frac{k \cdot q}{1-k \cdot q}} > 0.5.$$

Thus, normal agents are more likely to unilaterally defect than committed agents. One can show, that when there is a mutual defection, it is most likely that at least one of the agents involved is committed. This implies, that the partner is more likely to defect when observing him to be involved in mutual defection relative to observing him to be the sole defector. This implies that defection is the unique best reply when observing a single mutual defection, and this contradicts the assumption that normal agents cooperate with positive probability when observing a single mutual defection. When the game is mild, a construction similar to the previous proofs supports cooperation as a perfect equilibrium. □

Our last result studies the observation of actions against cooperation, and it shows that cooperation is a perfect equilibrium action in any underlying Prisoner's Dilemma. Formally:

Theorem 5. *Let $E = (G, p)$ be an environment with observation of actions against cooperation, where G is a PD game and $p \equiv k \geq 2$. Then cooperation is a perfect equilibrium action.*

The intuition behind the proof is the following. Not allowing Alice to observe Bob's behavior when his past opponent defected helps to sustain cooperation because it implies that defecting against a defector does not have any negative indirect effect (in any steady state) because it is never observed by future opponents.

This encourages agents to defect against partners who are more likely to defect, and allows cooperation to be sustained regardless of the values of g and l.

Remark 4. In the last two results (Theorems 4–5) cooperation is not a strictly perfect equilibrium. Specifically, it is not a perfect equilibrium action with respect to distributions of commitments in which the committed agents defect with high probability. The reason is that committed agents who defect with high probability induce normal partners to defect against them with probability one. This implies that when observing a partner to be involved on either side of a unilateral defection (either as the sole defector or as the sole cooperator), the partner is most likely to be normal. As a result the agents' incentives to defect are the same when observing mutual cooperation as when observing unilateral defection, and this does not allow cooperation to be supported in a perfect equilibrium, as such cooperation relies on agents who have better incentives to defect when observing a unilateral defection.

6 Discussion

6.1 Cheap Talk, Secret Handshakes and Equilibrium Selection

In this section we discuss two related issues: (1) in many setups both defection and cooperation are perfect equilibria; which of these equilibria is more likely to be selected? (2) What would be the influence of the introduction of pre-play "cheap-talk" communication to our setup?

For concreteness, we focus on the main model: observation of actions. As in the standard setup of normal-form games (without observation of past actions), the introduction of cheap talk induces different equilibrium selection results, depending whether or not deviators have unused messages to use as secret handshakes (see, e.g., Robson, 1990; Schlag, 1993; Kim and Sobel, 1995). If one assumes that the set of cheap-talk messages is finite, and all messages are costless, then cheap talk has little effect on the set of stable outcomes (as any perfect equilibrium of the game without cheap talk can be implemented as an equilibrium with cheap talk in which the incumbents send all messages with a positive probability).

In what follows we focus on a different case, in which there are slightly costly messages that, due to their positive cost, are not used unless they yield a benefit. In this setup our results should be adapted as follows.

1. Offensive games: No stable state exists. Both defection and cooperation are only "quasi-stable"; the population state occasionally changes between theses two states, based on the occurrence of rare random experimentations. The argument is adapted from Wiseman and Yilankaya (2001).

2. Defensive games (and $k \geq 2$): The introduction of cheap talk destabilizes all non-efficient equilibria leaving cooperation as the unique stable outcome. The argument is adapted from Robson (1990).

In what follows we only briefly sketch the arguments for these results, since a formal presentation would be very lengthy, and the contribution is somewhat limited given that similar arguments have already been presented in the literature.

Following Wiseman and Yilankaya (2001), we modify the environment by endowing agents with the ability to send a slightly costly message ϕ (called the *secret handshake*). An agent has to pay a small cost c either to send ϕ to her partner or to observe whether the partner has sent ϕ to her. In addition, we still assume that each agent observes $k \geq 2$ past actions of the partner. Let ξ be the initial small frequency of a group of experimenting agents (called *mutants*) who deviate jointly. We assume that $O(\epsilon) \cdot O(\xi) < c < O(\xi)$, i.e., that the small cost of the secret handshake is smaller than the initial share of mutants, but larger than the product

of the two small shares of the mutants ($O(\xi)$) and the committed agents ($O(\epsilon)$). To simplify the analysis we also assume that the committed agents do not use the secret handshake

Consider a population that starts at the defection equilibrium, in which all normal agents defect regardless of the observed actions and do not use signal ϕ. Consider a small group of ξ mutants ("cooperative handshakers") who send the signal ϕ, and cooperate iff the partner has sent ϕ as well. These mutants outperform the incumbents: they achieve ξ additional points by cooperating among themselves, which outweighs the cost of $2 \cdot c$ of using the secret handshake. Thus, assuming a payoff-monotonic selection dynamics, the mutants take over the population and destabilize the defective equilibrium. If the underlying game is offensive, then there is no other candidate to be a stable population state. Thus, cooperation can be sustained only until new mutants arrive ("defective handshakers") who use the secret handshake and always defect. These mutants outperform the cooperative handshakers, and would take over the population. Finally, a third group of mutants who always defect without using the secret handshake, can take the population back to the starting point.

If the underlying game is defensive, then there is a sequence of mutants who can take the population into the cooperative equilibrium characterized in the main text. Specifically, the second group of mutants (the ones after the cooperative handshakers) include agents who send only ϕ, but instead of incurring the small cost c of observing the partner's secret handshake, they base their behavior on the partner's observed actions, namely, they play some combination of the strategies s^1 and s^2. This second group of mutants would take over the population because the cost they save by not checking the secret handshake outweighs the small loss of $O(\epsilon)$ incurred from not defecting against committed partners. Finally, a third group of mutants who do not send the secret handshake, and follow strategies s^1 and s^2, can take over the population (by saving the cost of sending ϕ), and induce the perfect cooperative equilibrium of the main text. This equilibrium remains stable also with the option of using the secret handshake because: (1) mutants who defect when observing $m = 0$ are outperformed due to similar arguments to those in the main model, and (2) mutants who send the secret handshake, and always cooperate when observing ϕ (also when $m > 2$), are outperformed, as the cost of the secret handshake c outweighs the gain of $O(\xi) \cdot O(\epsilon)$.

6.2 Empirical Predictions and Experimental Verification

In this section we discuss a few testable empirical predictions of our model, and comment on how to evaluate these predictions in lab experiments.

An experimental setup to evaluate our predictions would include a large group of subjects (say, at least 10) who play a large number of rounds (say, on expectation at least 50 rounds), and are rematched at each period to play a Prisoner's Dilemma game with new partners. The experiment would include various treatments that differ in: (1) the parameters of the underlying game, e.g., if the game is offensive/defensive and mild/ acute; and (2) the information each agent observes about her partner. In particular, the number of past interactions that each agent observes, and what she observes in each interaction (e.g., actions, conflicts, or action profiles).

Our theoretical predictions deal with a "pure" setup in which all agents maximize their material payoffs except for a vanishingly small number of committed agents. An experimental setup (and, arguably, real-life interactions) differs in two key aspects: (1) agents, while caring about their material payoffs, may consider other non-material aspects, such as fairness and reciprocity; and (2) agents occasionally make mistakes; the frequency of these mistakes, while relatively low, is not negligible. In what follows, we describe our key predictions in the "pure" setup, interpret its implications in a "noisy" experimental setup, and describe the relevant existing data.

Our first prediction (Theorems 1–2) deals with observation of the partner's actions, and it states that cooperation can be sustained only in defensive games. In an experimental setup we interpret this to predict

that, *ceteris paribus*, the frequency of cooperation would be higher in a defensive game than in an offensive game. Engelmann and Fischbacher (2009), Molleman, van den Broek, and Egas (2013), and Swakman, Molleman, Ule, and Egas (2015) study the rate of cooperation in the borderline case of $g = l$ and in the closely related donor-recipient game, in which at each interaction only one of the players (the donor) chooses whether to give up g of her own payoff to yield a gain of $1 + g$ for the recipient. The typical findings in these experiments is that observation of 3–6 past actions induces a relatively high level of cooperation (50%–75%, where higher rates of cooperation are typically associated with environments in which more past actions are observed, or when subjects can also observe second order information about the behavior of the partner's past opponent). We are aware of only a single experiment that studies a setup in which $g \neq l$. Gong and Yang (2014) study the case of $g = 0.8 > l = 0.4$, and present results that are consistent with our prediction. They show that even though in their setup players observe 10 past actions of the partner, and, in addition, they are also able to observe second order information, the average rate of cooperation is only 30%–50%.

Our second prediction (Theorems 3–4) deals with observation of either past conflicts or past action profiles, and it states that cooperation can be sustained only in mild games. In an experimental setup it implies that, *ceteris paribus*, the frequency of cooperation would be higher in mild games than in acute games. We are unaware of any existing experiential data with observation of either action profiles or conflicts.

It is interesting to compare our first two predictions to the comparative statics that has been recently developed for repeated Prisoner's Dilemma games played by the same pair of players. Blonski, Ockenfels, and Spagnolo (2011), Dal Bó and Fréchette (2011), and Breitmoser (2015) present theoretical arguments and experimental data to suggest that when a pair of player repeatedly play the Prisoner's Dilemma, then the lower the values of g and l are, the easier it is to sustain cooperation.[28] However, our prediction is that when agents are randomly matched at each round, then the lower the value of g, and the *higher* the value of l, the easier it is to sustain cooperation.

Our final prediction is that when communities succeed in sustaining cooperation, it will be supported by the following behavior: most subjects defect (resp., cooperate, mix) when observing 2+ (resp., 0, 1) defections/conflicts. In an experimental setup we interpret this to predict that the probability that an agent defects increases with the number of times she observes the partner to be involved in defections/conflict. In particular, we predict a substantial increase in a subject's propensity to defect when moving from zero to two observations of defection. The findings of Engelmann and Fischbacher (2009), Molleman, van den Broek, and Egas (2013), Gong and Yang (2014), and Swakman, Molleman, Ule, and Egas (2015) suggest that subjects are indeed more likely to defect when they observe the partner to defect more often in the past.

6.3 Robustness of Results

In this section we discuss the robustness of our results with respect to various factors.

Finitely-Lived Agents, Non-stationary Strategies, and Observation of the k Last Actions The main model includes a few simplifying assumptions: (1) agents live forever and do not discount the future, (2) agents are only allowed to follow stationary strategies, (3) agents observe the partner's actions sampled from the entire infinite history of play of the partner, and (4) an agent does not know the actions observed by her partner about the agent's behavior. In Appendix A we relax all these assumptions, and show that cooperation

[28] Specifically, the above papers show that cooperation is more likely to be sustained in the infinitely repeated Prisoner's Dilemma if the discount factor of the players is above $\frac{g+l}{g+l+1}$. Note that this minimal threshold for cooperation is increasing in both parameters. Embrey, Frechette, and Yuksel (2015) present similar comparative statics evidence for the finitely repeated Prisoner's Dilemma.

remains a perfect equilibrium action in defensive Prisoner's Dilemma games, and that the behavior that sustains cooperation is similar to the main model.

Specifically, the model in Appendix A assumes that at each round a small share of $1-\delta$ of the agents are chosen at random to stop interacting (retire) and are replaced with new agents, who begin interacting with an empty history. When two agents interact, each of them observes the last k actions played by her partner (or the entire partner's history, if the partner has interacted in fewer than k rounds). Agents are allowed to follow any strategy (i.e., the strategy may depend on their personal history, and not only on the signal about the partner).

We adapt the definitions of steady state, Nash equilibrium, and perfect equilibrium, and we show that if cooperation is a perfect equilibrium action in any defensive Prisoner's Dilemma if the players observe at least two actions, and if δ is sufficiently close to one. The behavior that supports cooperation is quite similar to the unique behavior in the main model, except that it depends also on the agent's own recent history. Most of the time normal agents have a "perfect recent history," which means that they cooperated in the last $k-1$ rounds. In this case, the agent behaves similarly to the agent in the main model: she cooperates if she observed the partner to have always cooperated in the past, defects if she observed the partner to have defected at least twice, and she mixes if she observed the partner to have defected in the last round and cooperated in the preceding $k-1$ rounds. If the agent does not have a "perfect recent history" (i.e., if she defected at least once in the last $k-1$ rounds), then she cooperates until her own history becomes perfect again.

Joint Deviations and Evolutionary Stability Our main solution concept (namely, perfect equilibrium) considers only deviations by a single agent (who has mass zero in the infinite population). A stronger solution concept is the notion of evolutionarily stable strategy (Maynard Smith and Price, 1973) that requires also stability against a group of agents with a positive small mass who jointly deviate. This stronger notion also implies asymptotic stability in many monotonic dynamics (see, e.g., Sandholm, 2010); that is, if due to a small perturbation the population state slightly moves away from an evolutionarily stable state, then plausible dynamics, in which more successful strategies become more frequent, will take the population back to the state.

In Appendix B we adapt the notion of an evolutionarily stable strategy to the setup of an environment with observation of the partner's past behavior, and we show that the perfect equilibria that support cooperation in our main results satisfy the refinement of evolutionary stability.

General Noise Structures In the model described above we deal with perturbed environments that include a single kind of noise, namely, committed agents who follow commitment strategies. In what follows we discuss how to extend our results to include additional sources of noise: specifically, observation noise and/or trembles. We redefine a perturbed environment as a tuple $E_{\epsilon,\delta} = ((G,k),(S_C,\lambda),\alpha,\epsilon,\delta)$, where $(G,k),(S_C,\lambda),\epsilon$ are defined as in the main model, $0<\delta<<1$ is the probability of error in each observed action of a player, and $\alpha \in \Delta(A)$ is a totally mixed distribution according to which the observed error is sampled from in the event of an observation error. Alternatively, these errors can also be interpreted as action played by mistake by the partner due to trembling hands. The notion of a steady state (and, in particular, the consistent behavior) can be adapted to the setup with observation errors in a straightforward way. Indeed, one can show that *all* of our results can be adapted to this setup in a relatively straightforward way. In particular, our results hold also in environments in which most of the noise is due to observation errors, provided that there is a small positive share of committed agents (possibly much smaller than the probability of an observation error).[29]

[29]Formally, one needs to redefine a perfect equilibrium as the limit of Nash equilibria in a converging sequence of perturbed environments $((G,k),(S_C,\lambda),\alpha,\epsilon_n,\delta_n)$ where $\epsilon_n,\delta_n \to 0$. Next, we say that action $a \in A$ is a strictly perfect equilibrium in this extended setup if for any observed action converging sequence of perturbed environments $((G,k),(S_C,\lambda),\alpha,\epsilon_n,\delta_n)$ satisfying

7 Conclusion and Directions for Future Research

In many situations people engage in short-term interactions but future partners may obtain some information about the behavior today. The conventional approach to studying such environments assumes that all agents exactly follow the equilibrium behavior (and the source of any off-equilibrium path behavior is "trembling hands"), and that the entire community starts to interact at some hypothetical time zero. A stylized main result in the existing literature is that by relying on these two assumptions one can obtain various general folk theorem results. Often, these equilibria rely on agents playing complex strategies that depend on the exact calendar time (e.g., Kandori, 1992; Ellison, 1994; Takahashi, 2010).

We propose a new modeling approach in which (1) an equilibrium has to be robust to the presence of a few "crazy" agents, and (2) the community has been interacting from time immemorial. We develop a novel methodology that allows a tractable analysis of these seemingly complicated environments. We apply this methodology to study underlying Prisoner's Dilemma games (and coordination games), and we obtain sharp testable predictions for the equilibrium actions, including a striking anti-folk theorem result for offensive games. Finally, we show that whenever cooperation is sustainable, there is a unique (and novel) way to support it that has a few appealing properties: (1) agents behave in an intuitive and simple way, and (2) the equilibrium is robust to various perturbations (such as a few agents deviating together and the presence of crazy agents).

We believe that our novel modeling approach will be helpful in understanding many other interactions in future research. In particular, we plan to extend the methodology to asymmetric games, and to use it to study trade interactions between a population of consumers and professional sellers. Another direction for future research is to adapt the model to better fit online interactions, and to deal with non-verifiable public reports similar to the online feedback mechanisms in web-sites such as eBay. Finally, readers may be interested in our companion paper (Heller and Mohlin, 2015a), in which we study a related setup in which agents are allowed to exert effort in deception by influencing the message observed by the opponent.

References

Acemoglu, D., and A. Wolitzky (2014): "Cycles of conflict: An economic model," *The American Economic Review*, 104(4), 1350–1367.

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.

Bhaskar, V. (1998): "Informational constraints and the overlapping generations model: Folk and anti-folk theorems," *The Review of Economic Studies*, 65(1), 135–149.

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.

$\epsilon_n, \delta_n \to 0$ and $\frac{\epsilon_n}{\delta_n} \to$ constant (which is allowed to be 0 or ∞), there exists a converging sequence of Nash equilibria $(S_n, \sigma_n, \eta_n) \to (S^*, \sigma^*, \eta^* \equiv a)$ such that their outcomes converge to an outcome in which all normal agents play action a with probability one.

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

Cooper, B., and C. Wallace (2004): "Group selection and the evolution of altruism," *Oxford Economic Papers*, 56(2), 307–330.

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.

Deb, J. (2012): "Cooperation and community responsibility: A folk theorem for repeated matching games with names," *Available at SSRN 1213102*.

Deb, J., and J. González-Díaz (2014): "Community enforcement beyond the prisoner's dilemma," *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.

Eliaz, K., and A. Rubinstein (2014): "A model of boundedly rational "neuro" agents," *Economic Theory*, 57(3), 515–528.

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

Embrey, M., G. R. Frechette, and S. Yuksel (2015): "Cooperation in the Finitely Repeated Prisoner's Dilemma," *mimeo*.

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 (2014): "Reputation and cooperation: An experiment on prisoner dilemma with second-order information," *mimeo*.

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

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.

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.

Kim, Y.-G., and J. Sobel (1995): "An evolutionary approach to pre-play communication," *Econometrica*, pp. 1181–1193.

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.

Mailath, G. J., and L. Samuelson (2006): *Repeated games and reputations*, vol. 2. Oxford university press Oxford.

Matsushima, H., T. Tanaka, and T. Toyama (2013): "Behavioral Approach to Repeated Games with Private Monitoring," *CIRJE F-Series CIRJE-F-879, CIRJE, Faculty of Economics, University of Tokyo*.

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

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.

Molleman, L., E. van den Broek, and M. Egas (2013): "Personal experience and reputation interact in human decisions to help reciprocally," *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1757), 20123044.

Nowak, M. A., and K. Sigmund (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.

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

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.

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

Rubinstein, A., and A. Wolinsky (1985): "Equilibrium in a Market with Sequential Bargaining," *Econometrica*, 53(5), 1133–1150.

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

Schlag, K. H. (1993): "Cheap Talk and Evolutionary Dynamics," Bonn Department of Economics Discussion Paper B-242.

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.

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

Swakman, V., L. Molleman, A. Ule, and M. Egas (2015): "Reputation-based cooperation: empirical evidence for behavioral strategies," *Evolution and Human Behavior*.

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.

Weibull, J. W. (1995): *Evolutionary game theory*. The MIT press.

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

Young, H. P. (1993): "The evolution of conventions," *Econometrica*, pp. 57–84.

A Finitely Lived Agents and Non-stationary Strategies (for Online Publication)

In this appendix we adapt our model to deal with finitely lived agents who may choose non-stationary strategies, and who observe the most recent actions of the partner.

A.1 Adaptations to the Model

Environment As before, we consider an infinite population (a continuum of mass one) in discrete time. We redefine an *environment* to be a triple (G, k, δ), where G is the underlying symmetric game, $k \in \mathbb{N}$ is the number of recent actions of the agent that are observed by her partner, and $\delta \in (0,1)$ is the probability that a currently active agent is active also in the next round. Since the population is infinite (by the law of large numbers) in each time period a fraction $1-\delta$ of the population dies and is replaced by "newborn" agents. This means that the population consists of agents of different lengths of histories, called *ages*. The shares of different ages have geometric distribution with parameter δ: there are $1-\delta$ newborn agents with no history of play (age zero), $\delta \cdot (1-\delta)$ agents with a history of length one, $\delta^2 \cdot (1-\delta)$ agents with a history of length two, and so on.

In each period the agents are randomly matched into pairs and, before playing, each agent observes the last $\max\{k, T\}$ past actions of her partner, where T denotes the age of the partner. Let $M = \cup_{0 \leq i \leq k} A^i$ denote the set of all possible signals. Let $l(m)$ denote the length of the signal $m = \left(a_1, ..., a_{l(m)}\right) \in M$, i.e., the number of actions in the sequence. We interpret $a_{l(m)}$ as the most recent action of the partner, $a_{l(m)-1}$ as the second most recent action, and so on. Note that the partner's age is $l(m)$ if $l(m) < k$, and is at least k otherwise.

Remark 5. Our assumption that each agent observes the last k actions of the partner is made only to simplify the notation and to allow a more concise appendix. All of our results can be adapted to a more general setup in which each agent observes k actions randomly sampled from the partner's last $n \geq k$ actions. The case of $n >> k$ is the one closest to the main model. We choose to focus on the opposite case of $n = k$ (i.e., observing the last k actions) in order to demonstrate the robustness of our results in the setup that is the "furthest" from the main model.

We let $\mathbb{N}$ denote the set of natural numbers, including zero. A (private) history of an agent of age $T \in \mathbb{N}$ is a tuple $h_T = \left(\left(m_t, a_t, b_t\right)_{0 \leq t \leq T-1}, m_T\right)$, where $m_t \in M$ is the signal observed by the agent when her age was t, $a_t \in A$ is the action played by the agent at age t, and $b_t \in A$ is the action played by the past partner when the agent was of age t. Finally m_T is the signal the agent has observed about her current partner. Let H_T denote the set of all possible histories of length T, and let $H = \cup_{T \in \mathbb{N}} H_T$ denote the set of all histories.

A *strategy* is a mapping $s : H \to \Delta(A)$ assigning a mixed action to each private history. We redefine S to denote the set of all such strategies. Note, that unlike in the main model we do not impose any restrictions on the set of feasible strategies. In particular, we allow agents to follow non-stationary strategies. A strategy is *totally mixed* if for each history $h_T \in H$ and each action $a \in A$ it is the case that $s_{h_T}(a) > 0$.

Perturbed Environment A perturbed environment is a tuple consisting of: (1) an environment, (2) a distribution λ over a set of commitment strategies S_C that including a totally mixed strategy, and (3) a number ϵ representing how many agents are committed to playing strategies in S_C (*committed agents*). The remaining $1-\epsilon$ of the agents can play any strategy in S (*normal agents*). Formally:

Definition 12. A *perturbed environment* is a tuple $E_\epsilon = \left((G, k, \delta), (S_C, \lambda), \epsilon\right)$, where (G, k, δ) is an environment, S_C is a non-empty finite set of strategies (called *commitment strategies*) which includes a totally mixed

strategy, $\lambda \in \Delta(S_C)$ is a distribution with full support over the commitment strategies, and $\epsilon \geq 0$ is the mass of committed agents in the population.

A steady state (S, σ, η) of a perturbed environment $((G, k, \delta), (S_C, \lambda), \epsilon)$ is a triple including: (1) the finite set of strategies S, which is interpreted as the strategies followed by the normal agents; (2) a distribution σ with full support over S, interpreted as the distribution of the normal strategies in the population; and (3) the function $\eta_{(s,T)}(s', T)'$ that describes the distribution of actions played by an incumbent agents who follows strategy $s \in S \cup S_C$ at age T, conditional on being matched with an agent who follows strategy $s' \in S \cup S_C$ and has age T'. This behavior has to be consistent with the strategies of the agents. Formally:

Definition 13. A *steady state* (abbr., *state*) of a perturbed environment $((G, k, \delta), (S_C, \lambda), \epsilon)$ is a triple (S, σ, η) where $S \subseteq \mathcal{S}$ is a finite set of strategies (called *normal strategies*), $\sigma \in \Delta(S)$ is a distribution with a full support over S, and $\eta : (S \cup S_C \times \mathbb{N}) \times (S \cup S_C \times \mathbb{N}) \to \triangle(A)$ is a mapping (called *consistent behavior*) that assigns to each pair of strategies and ages $s, s' \in S \cup S_C$ and $T, T' \in \mathbb{N}$ a mixed action $\eta_{(s,T)}(s', T') \in \Delta(A)$ and satisfies the consistency condition below. Let $\bar{\eta}_{s,T}$ be the average distribution of actions played by an agent of age T who follows strategy s against a random partner:

$$\forall a \in A, \quad \bar{\eta}_{s,T}(a) := \sum_{s' \in S \cup S_C} \left[((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')) \cdot \sum_{T' \in \mathbb{N}} (1-\delta) \cdot \delta^{T'} \cdot \eta_{s,T}(s', T')(a) \right].$$

Let $\nu_{s,T}$ denote the distribution of signals that is observed when meeting a partner of age $T \in \mathbb{N}$ who follows strategy $s \in S \cup S_C$: if $l(m) = \min(T, k) > 0$ then

$$\nu_{s,T}(a_1, ..., a_{l(m)}) = \prod_{1 \leq i \leq l(m)} \bar{\eta}_{s,T-(l(m)+1)+i}(a_i);$$

if $l(m) \neq \min(T, k)$ then $\nu_{s,T}(m) = 0$; if $l(m) = \min(T, k) = 0$ then $\nu_{s,0}$ assigns mass one to observing an empty message. Let $\mu_{s,T} \in \Delta(M \times A \times A)$ denote a distribution over tuples, such that $\mu_{s,T}(m_T, a, b)$ is the probability that an agent of age T who follows strategy s observes signal m_T about her (random) partner, the agent plays action a, and the partner plays action b. Formally:

$$\mu_{s,T}(m_T, a, b) = \sum_{s' \in S \cup S_C} \left[((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')) \cdot \left(\sum_{T' \in \mathbb{N}} (1-\delta) \cdot \delta^{T'} \cdot \nu_{s',T'}(m_T) \cdot s_m(a) \cdot \eta_{s',T'}(s, T)(b) \right) \right].$$

The *consistency requirement* that η has to satisfy is that for each $a \in A$, $s, s' \in S \cup S_C$, and $T, T' \in \mathbb{N}$:

$$\eta_{s,T}(s', T')(a) = \sum_{((m_t, a_t, b_t)_{t \leq T-1}, m_T) \in H_T} \left(\prod_{0 \leq t \leq T-1} \mu_{s,t}(m_t, a_t, b_t) \right) \cdot \nu_{s',T'}(m_T) \cdot s_{((m_t, a_t, b_t)_{t \leq T-1}, m_T)}(a).$$

The consistency requirement, although defined by somewhat complex equations, has the same interpretation and intuitive explanation as the one presented for the consistency requirement in the main model.

Remark 6. An environment differs from a standard repeated game also in a setup in which agents are finitely-lived. In both approaches each agent knows her personal calendar time (age), and can condition her behavior on this information. In the standard repeated game approach, each agent also knows the global calendar time (i.e., the exact time that has passed since the first ever interaction in the community), and can condition her

behavior on this information. However, we assume that interactions within the community have occurred from time immemorial, and no agent knows the global calendar time. Thus, as in the main model of this paper, the behavior is not uniquely determined by the distribution of strategies, and one has to include consistent behavior as part of the description of the steady state.

Expected Payoff and Equilibria In what follows we define the (ex-ante) expected payoff of an agent who follows strategy s and has a probability of $1-\delta$ of stopping interacting at the end of each round, given a steady state (S,σ,η) of a perturbed environment $((G,k,\delta),(S_C,\lambda),\epsilon)$. When $s \in S \cup S_C$ is an incumbent strategy, we define the payoff as follows:

$$\begin{aligned}\pi_s(S,\sigma,\eta) &= (1-\delta)\cdot\sum_{T\in\mathbb{N}}\delta^T\cdot\sum_{s'\in S\cup S_C}\Bigg[((1-\epsilon)\cdot\sigma(s')+\epsilon\cdot\lambda(s'))\cdot \\ &\quad\cdot\sum_{T'\in\mathbb{N}}\left[(1-\delta)\cdot\delta^{T'}\cdot\sum_{(a,a')\in A^2}\eta_{s,T}(s',T')(a)\cdot\eta_{s',T'}(s,T)(a')\cdot\pi(a,a')\right]\Bigg]. \qquad (15)\end{aligned}$$

Eq. (15), while quite long, is a relatively straightforward formula for the sum of the expected payoffs of an agent during her entire active life. The first element of $(1-\delta)$ is normalization, as is common in formulas for discounted sums. The probability of an agent being alive at period T is δ^T. Given that the agent is alive, her partner follows each strategy s' with probability $(1-\epsilon)\cdot\sigma(s')+\epsilon\cdot\lambda(s')$, and his age is T' with a probability of $(1-\delta)\cdot\delta^{T'}$. The final sum is the expected payoff in interactions between the agent of age T and a partner of age T' who follows strategy s'. Let $\pi(S,\sigma,\eta)$ be the average payoff of the *normal* agents: $\pi(S,\sigma,\eta)=\sum_{s\in S}\sigma(s)\cdot\pi_s(S,\sigma,\eta)$.

Now consider an agent (Alice) who deviates and plays a new strategy $\hat{s}\in\mathcal{S}\backslash S$. Alice's strategy determines her behavior against the incumbents. This determines the distribution of signals that are observed by the partners when being matched with Alice, and thus it determines the incumbents' play against Alice, and Alice's payoff $\pi_{\hat{s}}(S,\sigma,\eta)$, in the new unique steady state that emerges following her deviation. The formal definition is a straightforward adaptation of the analogous definition in the main model. The formal definition is omitted for brevity (since it is long and include complicated notation and some technicalities).

Nash and Perfect Equilibrium The definitions of Nash equilibrium and perfect equilibrium remains essentially the same as in the main model. Formally:

Definition 14. A steady state (S,σ,η) of the perturbed environment $((G,k,\delta),(S_C,\lambda),\epsilon)$ is a *Nash equilibrium* if for each strategy $s\in\mathcal{S}$: $\pi_s(S,\sigma,\eta)\leq\pi(S,\sigma,\eta)$.

Definition 15. A steady state (S^*,σ^*,η^*) of the environment (G,k,δ) is a *perfect equilibrium* if there exist a distribution of commitments (S_C,λ) and converging sequences $(S_n,\sigma_n,\eta_n)_n\to_{n\to\infty}(S^*,\sigma^*,\eta^*)$ and $(\epsilon_n>0)_n\to_{n\to\infty}0$, such that for each n, the state (S_n,σ_n,η_n) is a Nash equilibrium of the perturbed environment $((G,k,\delta),(S_C,\lambda),\epsilon_n)$. If $\eta^*\equiv a$, we say that action $a\in A$ is a *perfect* equilibrium action.

A.2 Stability of Cooperation

For brevity, we present a single formal result in this section, which adapts Theorem 2 to the current setup (though, our other results can also be adapted to this setup in a similar way). Specifically, Theorem 6 below

shows that cooperation is a perfect equilibrium action in any defensive game if players observe at least two actions and δ is sufficiently close to one. Formally:

Theorem 6. *Let G_{PD} be a defensive Prisoner's Dilemma ($g < l$). If $k \geq 2$, and $\delta^{k-1} > \frac{l}{l+1}$, then cooperation is a perfect equilibrium action in the environment (G_{PD}, k, δ).*

Proof. Parts of the proof that are analogous to the proof of Theorem 2 are omitted or presented briefly. Let $s^\alpha \equiv \alpha$ be a strategy that defects with an arbitrary probability $\alpha \in (0,1)$ regardless of the observed signal and the private history. Let $\left((G,k,\delta),\left(\{s^\alpha\},1_{s^\alpha}\right),\epsilon\right)$ be a perturbed environment in which all the committed agents follow strategy s^α.

We say that an agent has a *perfect recent history*, if the agent's age is at least $k-1$, and she has cooperated in all the last $k-1$ rounds. For each $q \in (0,1)$, let s^q be the strategy that induces an agent who follows it to:

1. defect with probability one if: (1) the agent has a perfect recent history, and (2) the partner is observed to either (I) have an age lower than k and to have defected at least once, or (II) have an age of at least k and to have defected at least twice;

2. defect with probability $q \in (0,1)$ if: (1) the agent has a perfect recent history, and (2) the partner is observed to have defected in the last round and to have cooperated in the previous $k-1$ rounds; and

3. cooperate with probability one in all other cases.

In other words, an s^q-agent always cooperates if her own recent history is not perfect . If her recent history is perfect she defects with probability one if she observes her partner either (I) to have defected at least once before the age of $k-1$, or (II) to have defected at least twice. Finally, if her recent history is perfect she defects with probability q whenever she observes her partner to have defected in the last round and to have cooperated in the previous $k-1$ rounds.

In what follows we will show that if $\epsilon \in (0,1)$ is sufficiently close to zero and $\delta \in (0,1)$ is sufficiently close to one, then there exists q_ϵ, such that $\left(\{s^{q_\epsilon}\},1_{s^{q_\epsilon}},\eta_\epsilon\right)$ is a Nash equilibrium of the perturbed environment $\left((G,k,\delta),\left(\{s^\alpha\},1_{s^\alpha}\right),\epsilon\right)$, and that $\left(\{s^{q_\epsilon}\},1_{s^{q_\epsilon}},\eta_\epsilon\right) \to_{\epsilon\to 0} \left(\{s_{q^*}\},1_{s_{q^*}},\eta^* \equiv c\right)$, which implies that cooperation is a perfect equilibrium action in the environment (G_{PD},k,δ).

Let $m_{k-1,1} = (c,...,c,d)$ denote a signal of length k that includes $k-1$ cooperations followed by a single defection, and let $m_{k,0}$ denote the signal that includes k cooperations. Let μ_ϵ be the posterior probability that the partner is going to defect conditional on the agent observing signal $m_{k-1,1}$, and the partner observing signal[30] $m_{k,0}$. Observe that $\mu_\epsilon \in (0,\alpha)$ is a decreasing function of q_ϵ: the larger the value of q_ϵ, the more likely that the partner is normal conditional on the agent observing signal $m_{k-1,1}$, and, thus, the more likely that the partner will cooperate conditional on the partner observing $m_{k,0}$ (as normal agents always cooperate when observing $m_{k,0}$).

Recall that an agent has a *perfect recent history*, if the agent's age is at least $k-1$, and she has cooperated in all the last $k-1$ rounds. Let $m_{k-1,1} = (c,...,c,d)$ denote a signal of length k that includes $k-1$ cooperations followed by a single defection, and let $m_{k,0}$ denote the signal that includes k cooperations. Let μ_ϵ be the posterior probability that the partner is going to defect conditional on the agent observing signal $m_{k-1,1}$, and the partner observing signal $m_{k,0}$. Observe that $\mu_\epsilon \in (0,\alpha)$ is a decreasing function of q_ϵ: the larger the value of q_ϵ, the more likely that the partner is normal conditional on the agent observing signal $m_{k-1,1}$, and, thus, the more likely that the partner will cooperate conditional on the partner observing $m_{k,0}$ (as normal agents always cooperate when observing $m_{k,0}$).

[30] The same probability is induced also in cases in which the partner observes a single defection followed by $k-1$ cooperations.

The value of q_ϵ is determined such that both actions are best replies when an agent with a perfect recent history observes signal $m_{k-1,1}$. Defection in such a case yields an immediate gain of $\mu_\epsilon \cdot l + (1-\mu_\epsilon) \cdot g$, and an indirect loss of $\delta \cdot q_\epsilon \cdot (l+1) + O(\epsilon)$ due to inducing the partner in the next round (if there is a next round and if the next partner is normal) to defect with probability q_ϵ instead of cooperating. This implies that q_ϵ has to solve the following equation:

$$\mu_\epsilon \cdot l + (1-\mu_\epsilon) \cdot g = \delta \cdot q_\epsilon \cdot (l+1) + O(\epsilon) \Rightarrow q_\epsilon = \frac{\mu_\epsilon \cdot l + (1-\mu_\epsilon) \cdot g}{\delta \cdot (l+1)} + O(\epsilon). \tag{16}$$

Observe that Eq. (16) admits a solution $q_\epsilon(\mu_\epsilon) \in (0,1)$ for any value of μ_ϵ as long as $\delta > \frac{l}{l+1}$. Moreover, this solution is increasing in μ_ϵ. This implies (together with the above observation that $\mu_\epsilon(q_\epsilon)$ is decreasing in q_ϵ) that for any sufficiently small $\epsilon > 0$ there is a unique pair $(q_\epsilon, \mu_\epsilon)$ that solves Eq. (16), and satisfies the condition that μ_ϵ is induced by q_ϵ via Bayes' rule. Furthermore, as $\epsilon \to 0$, the pair converges to some (q^*, μ^*).

Next, observe that for any sufficiently small ϵ there indeed exists a consistent behavior η_ϵ, such that normal agents cooperate with high probability among themselves (i.e., for each $T, T' \in \mathbb{N}$, $(\eta_\epsilon)_{(s,T)}(s', T') = 1 - O(\epsilon)$). The argument for this can be sketched as follows (the argument is presented in brief as it is similar to the one in the proof of Theorem 2). The committed agents defect with a probability of $O(\epsilon)$. Normal agents that are matched with committed agents are likely to defect due to observing the committed partner defecting in the past. This induces $O(\epsilon)$ defections by normal agents. These defections induce additional $q_\epsilon \cdot O(\epsilon) + O(\epsilon^2)$ defections by normal agents who are being matched with normal partners. These additional defections induce a further $q_\epsilon \cdot \left(q_\epsilon \cdot O(\epsilon) + O(\epsilon^2)\right) + O(\epsilon^2)$ defections, and so on. One can show that this process implies that the total average probability of defections when two normal agents are matched is $\frac{O(\epsilon)}{1-q_\epsilon} + O(\epsilon^2) = O(\epsilon)$. Thus $\eta_\epsilon \to_{\epsilon \to 0} \eta^* \equiv c$.

Analogous arguments to the one presented in the proof of Theorem 2 imply that when an agent with a perfect recent history observes the partner to be always cooperative, then the partner is most likely to be normal and to cooperate against the agent, and the agent's unique best reply is to cooperate. This is because the direct gain from defecting is smaller if the partner is cooperative because the game is defensive, while the indirect loss from defecting is independent of the partner's behavior. Similarly, when an agent with a perfect recent history observes the partner to defect at least twice out of k observations, or at least once out of a smaller sample, then the partner's posterior probability of defecting against the agent is larger than μ_ϵ, and the partner's unique best reply is to defect.

Finally, the unique best reply of an agent who does not have a perfect recent history is to cooperate till his recent history becomes perfect. Defecting while having an "imperfect" recent history induces normal partners to defect for sure, and yields the agent a payoff of at most 0 in each round. If, however, an agent cooperates until his recent history becomes perfect, he will get at least $-l$ for at most $k-2$ rounds (until her recent history becomes perfect again), and then she is likely to get $1 - O(\epsilon)$ in all future rounds. Thus cooperating till the recent history becomes perfect is the unique best reply as long as δ is sufficiently high (neglecting terms of $O(\epsilon)$):

$$0 < (-l) \cdot \left(1 + \delta + \delta^2 + ... + \delta^{k-2}\right) + 1 \cdot \left(\delta^{k-1} + \delta^k + ...\right) \Leftrightarrow \frac{l \cdot \left(1 - \delta^{k-1}\right)}{1-\delta} < \frac{\delta^{k-1}}{1-\delta} \Leftrightarrow \delta^{k-1} > \frac{l}{l+1}.$$

The above arguments show that $\left(\{s^{q_\epsilon}\}, 1_{s^{q_\epsilon}}, \eta_\epsilon\right)$ is indeed a Nash equilibrium, and thus $\left(\{s_{q^*}\}, 1_{s_{q^*}}, \eta^* \equiv c\right)$ is a perfect equilibrium. □

B Evolutionary Stability (for Online Publication)

In the main text we have dealt with the notion of perfect equilibrium, which requires that no agent be able to achieve a better payoff than the incumbents by unilateral deviation. In this appendix we refine the solution concept to require stability also against small groups of agents (with a positive small mass) who deviate together. It turns out that all of our results also work under this refinement.

B.1 Definitions

In a seminal paper, Maynard Smith and Price (1973) define a symmetric Nash equilibrium strategy α^* to be evolutionarily stable if the incumbents achieve a strictly higher payoff when being matched with any other best reply strategy β (i.e., $\pi(\beta,\alpha^*)=\pi(\alpha^*,\alpha^*) \Rightarrow \pi(\alpha^*,\beta)>\pi(\beta,\beta)$. The motivation is that if β is a best reply to α^*, then a single deviator who plays β will be as successful as the incumbents. This may induce a few other agents to mimic her behavior, until a small positive mass of agents follow β . The above inequality implies that at this stage the followers of β will be strictly outperformed, and thus will disappear from the population.

Our setup with environments is similar to the standard setup of a repeated game in that it rarely admits evolutionarily stable strategies. Typically, not all the actions will be played by normal agents in equilibrium, and as a result some signals will never be observed. Deviators who differ in their behavior only after such zero probability signals will get the same payoff as the incumbents both against the incumbents and against other deviators. This violates the above inequality.

Following Selten's (1983) notion of "limit ESS," we solve this issue by requiring evolutionary stability in a converging sequence of perturbed environments, in which all signals are observed on the equilibrium path, instead of simply requiring evolutionary stability in the unperturbed environment.

This is formalized as follows. Given a steady state (S,σ,η) in a perturbed environment $((G,k),(S_C,\lambda),\epsilon)$, we define $\pi_{\hat{s}}(\hat{s})$ as the (long-run average) payoff of strategy $\hat{s}$ against itself, and $\pi_{(S,\sigma)}(\hat{s})$ as the mean (long-run average) payoff of the incumbents against strategy $\hat{s}$. Specifically, if $\hat{s}\in S\cup S_C$, then

$$\pi_{\hat{s}}(\hat{s}|S,\sigma,\eta)=\sum_{(a,a')\in A^2}\hat{\eta}_{\hat{s}}(\hat{s})(a)\cdot\hat{\eta}_{\hat{s}}(\hat{s})(a')\cdot\pi(a,a'),$$

$$\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\eta)=\sum_{s\in S\cup S_C}\sum_{(a,a')\in A^2}((1-\epsilon)\cdot\sigma(s)+\epsilon\cdot\lambda(s))\cdot\eta_s(\hat{s})(a)\cdot\hat{\eta}_{\hat{s}}(s)(a')\cdot\pi(a,a'),$$

and if $\hat{s}\notin S\cup S_C$, then we define $\pi_{\hat{s}}(\hat{s})$ and $\pi_{(S,\sigma)}(\hat{s})$ as the respective payoffs in the post-deviation steady state $(S\cup\{\hat{s}\},\hat{\sigma},\hat{\eta})$:

$$\pi_{\hat{s}}(\hat{s}|S,\sigma,\eta)=\sum_{(a,a')\in A^2}\hat{\eta}_{\hat{s}}(\hat{s})(a)\cdot\hat{\eta}_{\hat{s}}(\hat{s})(a')\cdot\pi(a,a'),$$

$$\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\eta)=\sum_{s\in S\cup S_C}\sum_{(a,a')\in A^2}((1-\epsilon)\cdot\hat{\sigma}(s)+\epsilon\cdot\lambda(s))\cdot\hat{\eta}_s(\hat{s})(a)\cdot\hat{\eta}_{\hat{s}}(s)(a')\cdot\pi(a,a').$$

Definition 16. A steady state (S^*,σ^*,η^*) of a perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ is *evolutionarily stable* if: (1) (S^*,σ^*,η^*) is a Nash equilibrium, and (2) for any best reply strategy $\hat{s}$ (i.e., $\pi_{\hat{s}}(S^*,\sigma^*,\eta^*)=\pi(S^*,\sigma^*,\eta^*)$), such that $\sigma^*(\hat{s})<1$ (i.e., $\hat{s}$ is not the only normal strategy) the following inequality holds: $\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\eta)>\pi_{\hat{s}}(\hat{s}|S,\sigma,\eta)$.

Definition 17. A steady state (S^*,σ^*,η^*) of the environment (G,k) is *a perfect evolutionarily stable state* if there exist a distribution of commitments (S_C,λ) and converging sequences $(S_n,\sigma_n,\eta_n)_n\to_{n\to\infty}(S^*,\sigma^*,\eta^*)$

and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state (S_n, σ_n, η_n) is an evolutionarily stable state in the perturbed environment $((G,k),(S_C,\lambda),\epsilon_n)$. If the outcome assigns probability one to one of the actions, i.e., $\eta^* \equiv a$, then we say that this action is a perfect evolutionarily stable outcome.

Finally, we define a strictly perfect evolutionarily stable outcome as a pure action that is an outcome of a perfect evolutionarily stable state for any distribution of commitments (similar to the notion of strict limit ESS in Heller, 2015b).

Definition 18. Action $a^* \in A$ is a *strictly perfect evolutionarily stable outcome* in the environment $E = ((A,\pi),k)$ if, for any distribution of commitment strategies (S_C,λ), there exist a steady state $(S^*,\sigma^*,\eta^* \equiv a^*)$ and converging sequences $(S_n,\sigma_n,\eta_n)_n \to_{n\to\infty} (S^*,\sigma^*,\eta^*)$ and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state (S_n,σ_n,η_n) is an evolutionarily stable state in the perturbed environment $((G,k),(S_C,\lambda),\epsilon_n)$.

B.2 Adaptation of Results

All of our results holds with respect to the refinement of evolutionary stability. In particular, the fact that always defecting is a strict equilibrium (i.e., the unique best reply to itself) in any slightly perturbed environment, implies that defection is a strictly perfect evolutionarily stable outcome.

Similarly, one can adapt the results about sustaining cooperation as an equilibrium action (Theorems 2–5). Specifically, minor modifications to the proofs can show that cooperation is a strictly perfect evolutionarily stable outcome in defensive games with observation actions and in mild games with observation of conflicts (when $k \geq 2$), and that cooperation is a perfect evolutionarily stable outcome in mild games with observation of action profiles, and in any game with observation of actions against defectors.

A sketch of the argument why the results apply also to the refinement of evolutionary stability is as follows. There are two kinds of steady states that sustain cooperation in the proofs in this paper:

1. Steady state $\psi'_n = (s^{q_n}, 1_{s^{q_n}}, \eta_n)$ that has a single normal strategy in its support. The arguments in the proofs show that each such strategy is the unique best reply to itself in the n^{th} perturbed environment (i.e., $\pi_{s'}(\psi'_n) < \pi(\psi'_n)$ for each $s' \neq s^{q_n}$), which shows that ψ'_n is an evolutionarily stable state in the n^{th} perturbed environment.

2. Steady state $\psi_n = \left(\left\{s^1, s^2\right\}, (q_n, 1-q_n), \eta_n\right)$ that consists of two normal strategies in its support. The arguments in the proofs show that these two strategies are the only best replies to this steady state (i.e., $\pi_{s'}(\psi_n) < \pi(\psi_n\eta_n)$ for each $s' \notin \left\{s^1, s^2\right\}$). Moreover, the arguments in the proof (see, in particular, Remark 8 at the end of the proof of Theorem 2) imply that each of these two normal strategies obtains a relatively low payoff when being matched against itself, i.e.,: $\pi\left(s^1|\psi_n\right) > \pi_{s^1}\left(s^1|S,\sigma,\eta\right)$ and $\pi\left(s^2|\psi_n\right) > \pi_{s^2}\left(s^2|\psi_n\right)$, which implies that ψ_n is evolutionarily stable.

C Proofs (for Online Publication)

C.1 Proof of Lemma 1 (Existence of Consistent Behavior)

Fix environment $(G=(A,\pi),k)$, a finite set of strategies S, and a distribution $\sigma\in\Delta(S)$. Let O_S be the set of all behaviors defined over S, i.e., the set of all mappings of the form $\eta: S\times S\to\Delta(A)$. Let $f_\sigma: O_S\to O_S$ be the transformation between behavior mappings that is induced by σ. That is, $f_\sigma(\eta)$ is the "new" behavior that is induced by players who follow strategy distribution σ, and observe signals about the partners according to the "old" behavior η. Formally:

$$\forall a\in A,\ s,s'\in S,\quad \left(f_\sigma(\eta)\right)_s(s')(a)=s_{\nu_{\bar{\eta}_{s'}}}.$$

Observe that the space O_S is a convex and compact subset of a Euclidean space, and that the mapping $f_\sigma: O_S\to O_S$ is continuous. Brouwer's fixed-point theorem implies that the mapping σ has a fixed point, which is a consistent outcome by definition.

C.2 Proof of Proposition 1 (Implementation of Perfect Equilibria)

If α is a totally mixed strategy, then it is immediate that the state $(\{\alpha\},1_\alpha,\alpha)$ is a Nash equilibrium of the perturbed environment$((G,k),(\{\alpha\},1_\alpha),\epsilon)$ for any $\epsilon>0$, which implies that the state $(\{\alpha\},1_\alpha,\alpha)$ is a perfect equilibrium. Assume now that α is not totally mixed. The fact that $\alpha\in\Delta(A)$ is a symmetric perfect equilibrium of the underlying game implies (see Selten, 1975, Theorem 7) that there is a sequence of totally mixed strategies $(\alpha_n)\to_{n\to\infty}\alpha$, such that α is a best reply to each α_n. The fact that α is a best reply both to itself and to α_1 (the first element in the sequence (α_n)) implies that the state $(\{\alpha\},1_\alpha,\alpha)$ is a Nash equilibrium of the regular perturbed environment $((G,k),(\{\alpha_1,\alpha\},(0.5,0.5)),\epsilon)$ for any $\epsilon>0$, which implies that $(\{\alpha\},1_\alpha,\alpha)$ is a regular perfect equilibrium.

C.3 Proof of Proposition 2 (Mixed Equilibrium in Coordination Game)

Assume to the contrary that $(\{\alpha\},1_\alpha,\alpha)$ is a regular perfect equilibrium in the environment $(G,k\geq 1)$. This implies that $(\{\alpha\},1_\alpha,\alpha)$ is a Nash equilibrium of some regular perturbed environment $((G,k),(S_C,\lambda),\epsilon>0)$. The regularity of (S_C,λ) implies that there is $s\in S_C$ such that $s_{\nu_\alpha}\neq\alpha$. Assume w.l.o.g. that $s_{\nu_\alpha}(a)>\alpha(a)$. This inequality implies that when an agent observes a signal $m_{\vec{a}}=(a,...,a)$ (i.e., the partner played the action a in all k observed interactions), then there is a posterior probability strictly larger than $\alpha(a)$ that the partner is going to play a. This implies that playing a when observing signal $m_{\vec{a}}$ induces a strictly larger payoff than playing α, which contradicts $(\{\alpha\},1_\alpha,\alpha)$ being a Nash equilibrium in the regular perturbed environment.

C.4 Proof of Proposition 3 (Strictly Perfect Outcomes in Coordination Games)

We identify each signal m with the number of times that action b has been played in the sample of k observations.

Case I: Suppose $\pi(a,a)<\pi(b,b)$. We want to show that a is not a strictly perfect equilibrium action. Assume to the contrary that a is a strictly perfect equilibrium action. Let s^α be the strategy such that $s^\alpha_k(a)=\alpha$ (and $s^\alpha_k(b)=1-\alpha$) for all k. Pick $\alpha>0$ sufficiently small such that b is the unique best reply against s^α. Consider a perturbed environment $((G,k),(\{s^\alpha\},1_{s^\alpha}),\epsilon>0)$. The assumption that a is strictly perfect implies that there is a steady state $(S^*,\sigma^*,\eta^*\equiv a)$, a converging sequence of steady states $(S_n,\sigma_n,\eta_n)\to(S^*,\sigma^*,\eta^*)$,

and a converging sequence of perturbed environments $((G,k),(\{s^\alpha\},1_{s^\alpha}),\epsilon_n)$, such that each (S_n,σ_n,η_n) is a Nash equilibrium of $((G,k),(\{s^\alpha\},1_{s^\alpha}),\epsilon_n)$. Fix a sufficiently small ϵ_n (sufficiently large n).

Assume first that $(s_n)_k(b)=1$ for each $s_n\in S_n$ (i.e., all normal agents play b with probability one if they observe only b's). This implies that a deviating agent (Alice) who always plays b outperforms the incumbents: Alice will get a high payoff very close to $\pi(b,b)$ (because both she and all of her normal partners play b), while the incumbents achieve a lower average payoff of about $\pi(a,a)$ (because $\eta_n\to_{n\to\infty}\eta^*\equiv a$). This contradicts (S_n,σ_n,η_n) being a Nash equilibrium of $((G,k),(\{s^\alpha\},1_{s^\alpha}),\epsilon_n)$.

Next assume that there is a strategy $s_n\in S_n$ such that $(s_n)_k(b)<1$. Note that when an agent observes signal $m=k$, it implies that with high probability the partner is following the commitment strategy s^α (because $\eta_n\to_{n\to\infty}\eta^*\equiv a$) , so that the unique "myopic" best reply (taking into account only the payoff in this interaction, and not the fact that the action is observed by future partners) is action b. The fact that a normal agent who follows s_n plays a with positive probability when she observes signal $m=k$ implies that the direct loss of playing a when observing k must be compensated by the indirect gain accruing from interactions with future partners who observe the current interaction (otherwise (S_n,σ_n,η_n) could not be a Nash equilibrium). This indirect future gain is independent of the current partner's behavior, while the direct loss from playing a is strictly larger when observing $m=k$ than when observing $m<k$. Hence, playing a is the unique best reply when an agent observes any signal $m<k$ (taking into account both the direct and the indirect impact of the played action on the payoff). In particular, all normal agents play a when observing $m=1$. This implies that the indirect loss of playing b when observing $m=k$ is very small $(O(\epsilon_n^k))$ because the probability of observing the signal $m=k$ is small $(O(\epsilon_n))$, and hence it is very unlikely $(O(\epsilon_n^k))$ that a future opponent will observe only interactions in which the agent played b because she observed the signal $m=k$. Thus the indirect gain of playing b when observing $m=k$ (which is $O(\epsilon_n)$) strictly outweighs the indirect loss (which is $O(\epsilon_n^k)$) and thus b is the unique best reply when an agent observes $m=k$. Hence (S_n,σ_n,η_n) cannot be a Nash equilibrium if there is a strategy $s_n\in S_n$ such that $(s_n)_k(b)<1$.

Case II: Suppose $\pi(a,a)>\pi(b,b)$. We wish to show that a is a strictly perfect equilibrium action. Let (S_C,λ) be an arbitrary distribution of commitments. Let $\bar\beta\in(0,1)$ be the probability of action b in the unique mixed equilibrium of the underlying game G. Let $(\lambda|m)\in\Delta(S_C)$ be the posterior distribution of the partner's strategy, conditional about the partner following a commitment strategy, and the agent observing signal m on the partner, in a population in which everyone observes the message $m=0$ (which is the relevant case since we need $\eta_n\to_{n\to\infty}\eta^*\equiv a$ in order for a to be a strictly perfect equilibrium action). Formally (by using Bayes' rule):

$$(\lambda|m)(s)=\frac{\lambda(s)\cdot\nu_{s_0}(m)}{\sum_{s\in S_C}\lambda(s)\cdot\nu_{s_0}(m)}.$$

Let $\beta_C(m)$ be the posterior probability that a random partner plays b conditional on (1) the agent observing signal m about the partner, (2) the partner following a commitment strategy, and (3) the partner observing signal 0 about the agent. Formally:

$$\beta_C(m)=\sum_{s\in S_C}(\lambda|m)(s)\cdot s_0(b).$$

It is straightforward to see that $\beta_C(m)$ is weakly increasing in m provided that ϵ is sufficiently small. (Note that if ϵ is very small then $s_0(b)$ is very close to the average probability that strategy s plays b.) Let $s^{\hat m}$ be the strategy that plays action a iff $m<\hat m$, i.e., $s_m^{\hat m}(a)=1$ if $m<\bar m$, and $s_m^{\hat m}(a)=0$ if $m\geq\hat m$.

Remark 7. In order to shorten the remaining proof, we take a simplifying assumption that for each m, $\beta_C(k)\neq$

$\bar{\beta}$. The knife-edge cases in which $\beta_C(m) = \bar{\beta}$ for some $m \in M$ complicates the proof, and makes it substantially longer, which, we felt is not justified given that the result is not the main focus of the paper.

To complete the proof (under the above simplifying assumption) we consider three exhaustive and mutually exclusive cases:

1. $\beta_C(k) < \bar{\beta}$. This implies that the steady state $(\{a\}, 1_a, \eta \equiv a)$, where η is any consistent behavior, is a Nash equilibrium of the perturbed environment $((G,k),(S_C,\lambda),\epsilon)$ for any sufficiently small $\epsilon > 0$.

2. $\beta_C(1) < \bar{\beta} \leq \beta_C(k)$. Let $\bar{m} > 1$ be the minimal signal m such that $\bar{\beta} < \beta_C(\bar{m})$. Then the steady state $(\{s^{\bar{m}}\}, 1_{s^{\bar{m}}}, \eta)$, where η is any consistent behavior in which the normal agents play action a with a high probability $(\eta_{s^{\bar{m}}}(s^{\bar{m}}) > 1 - O(\epsilon))$, is a Nash equilibrium of $((G,k),(S_C,\lambda),\epsilon)$ for any sufficiently small $\epsilon > 0$.

3. $\bar{\beta} < \beta_C(1)$. Let s^1 (s^2) be the strategy that induces an agent to play b iff $m \geq 1$ ($m \geq 2$). For each $q \in \left(0, \frac{1}{k}\right)$, consider the steady state $\left(\left\{s^1, s^2\right\}, (q, 1-q), \eta\right)$ of the perturbed environment $((G,k),(S_C,\lambda),\epsilon)$, where η is a consistent behavior in which the normal agents play a with an average probability of $1 - O(\epsilon)$ (such a consistent behavior exists due to the same arguments for the existence of a consistent behavior in which players cooperate with probability of $1 - O(\epsilon)$ in the proof of Theorem 2).
   Let μ_q be the posterior probability that a random partner is going to play b conditional on (1) the agent observing signal $m = 1$ about the partner, and (2) the partner observing signal $m = 0$ about the agent. Observe that (for a sufficiently small ϵ): (1) $\mu_0 = \beta_C(1) + O(\epsilon) > \bar{\beta}$, (2) μ_q is decreasing in q, and (3) $lim_{q \to \frac{1}{k}} \mu_q = O(\epsilon)$ (this is because each interaction in which a committed agent plays action b induces $O\left(\frac{1}{1-k \cdot q}\right)$ interactions in which normal agents play action b, as discussed in detail in the proof of Theorem 2, (see, in particular, Eq. (17)).
   This implies that for every sufficiently small ϵ there is a value of q_ϵ such that $\mu_{q_\epsilon} = \bar{\beta}$ (and that this value converges to $q_0 \in \left(0, \frac{1}{k}\right)$ as ϵ converges to zero. In the steady state $\left(\left\{s^1, s^2\right\}, (q_\epsilon, 1-q_\epsilon), \eta\right)$ both actions a and b are best replies conditional on observing signal $m = 1$, while action a (b) is the unique best reply when observing $m = 0$ ($m > 1$). This implies that $\left(\left\{s^1, s^2\right\}, (q_\epsilon, 1-q_\epsilon), \eta\right)$ is a Nash equilibrium of[31] $((G,k),(S_C,\lambda),\epsilon)$.

In all three cases we have characterized a converging sequence of Nash equilibria of the perturbed environments $((G,k),(S_C,\lambda),\epsilon_n)$ in which all the normal agents play action a with an average probability of $1 - O(\epsilon)$, which implies that action a is strictly perfect.

C.5 Proof of Proposition 4 (Defection is Strictly Perfect)

Let $\zeta = (\mathcal{S}_\mathcal{C}, \lambda)$ be a distribution of commitments. Let $s_d \equiv d$ be the strategy that always defects. Let $(\{s_d\}, 1_{s_d}, \eta_n)$ be a steady state of the perturbed environment $((G,k),(S_C,\lambda),\epsilon_n)$. Note that $(\eta_n)_{s_d}(s')(d) = 1$ for each strategy $s' \in S_C \cup \{s_d\}$. Consider a deviating agent (Alice) who follows any strategy $s \neq s_d$. We show that Alice is strictly outperformed in any post-deviation steady state.

The facts that $s \neq s_d$ and that all signals are observed with a positive probability in any perturbed environment imply that Alice cooperates with an average probability of $\alpha > 0$. We now compare the payoff of Alice

[31] We have abstracted away from a technical issue (which is formally investigated in the analogous arguments in the proof of Theorem 2). Specifically, we implicitly assumed that the probability that a random player defects conditional on both players observing $m = 1$ (denoted by the parameter $\chi \equiv \chi_{q_\epsilon}$ at the end of the proof of Theorem 2) is larger than μ_{q_ϵ}. Some distribution of commitment strategies might induce a situation in which $\chi_{q_\epsilon} < \mu_{q_\epsilon}$. In these cases, one needs to adapt the argument above by having the steady state $(\{s^{q_\epsilon}\}, 1_{s^{q_\epsilon}}, \eta)$, where s^{q_ϵ} is the strategy that plays b with probability q_ϵ when observing $m = 1$, plays a for sure when observing $m = 0$, and plays b for sure when observing $m > 1$.

to the payoff of an incumbent (Bob) who follows s_d. Alice obtains a direct loss of at least $\alpha \cdot \min(g, l)$ due to cooperating with probability α. The maximal indirect benefit that she might achieve due to these cooperations (by inducing committed agents to cooperate against her with higher probability relative to their cooperation probability against Bob) is $\epsilon_n \cdot k \cdot \alpha \cdot (l+1)$ because there are ϵ_n committed agents, each of whom observes Alice cooperate at least once in the k sampled actions with a probability of at most $k \cdot \alpha$, and each committed agent can yield Alice a benefit of at most $l+1$ by cooperating when he observes $m \geq 1$. If ϵ_n is sufficiently small ($\epsilon_n < \frac{1}{k \cdot (l+1)}$), then the direct loss is larger than the indirect maximal benefit ($\alpha > \epsilon_n \cdot k \cdot \alpha \cdot (l+1)$). This implies that $(\{s_d\}, 1_{s_d}, \eta_n)$ is a (strict) Nash equilibrium in any environment with $\epsilon_n < \frac{1}{k \cdot (l+1)}$, which proves defection is a strictly perfect equilibrium action.

C.6 Proof of Theorem 1 (Defection is the Unique Equilibrium in Offensive PDs)

Let (S^*, σ^*, η^*) be a regular perfect equilibrium. That is, there exists a regular distribution of commitments $(\mathcal{S}_C, \lambda)$, a converging sequence $(\epsilon_n)_n \to 0$, and a converging sequence of steady states $(S_n, \sigma_n, \eta_n) \to (S^*, \sigma^*, \eta^*)$, such that for each n the state (S_n, σ_n, η_n) is a Nash equilibrium of $((G,k), (S_C, \lambda), \epsilon_n)$. We assume to the contrary that $S^* \neq \{d\}$.

Recall that any message $m \in M = \{0, ..., k\}$ is observed with positive probability in any perturbed environment. Given a state (S_n, σ_n, η_n), an environment $((G,k), (S_C, \lambda), \epsilon_n)$, a message $m \in M$, and a strategy $s \in S_n$, let $q(m, s)$ denote the 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 19. Strategy $s \in S_n$ is *defector-favoring* given state (S_n, σ_n, η_n) and environment $((G,k), (S_C, \lambda), \epsilon_n)$ if there is some $\bar{q} \in [0,1]$ such that, for each $m \in M$, $q(m,s) > \bar{q} \Rightarrow s_m(d) = 0$, and $q(m,s) < \bar{q} \Rightarrow s_m(d) = 1$.

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

First, we show that all normal strategies are defector-favoring. Assume to the contrary that there is a strategy $s \in S_n$ that is not defector-favoring. Let s' be a defector-favoring strategy that has the same average defection probability as s in the post-deviation steady state. 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'), and 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 underlying game is offensive this implies that strategy s' strictly outperforms strategy s, which contradicts (S_n, σ_n, η_n) being a Nash equilibrium.

Second, we show that all the normal strategies defect with the same average probability in (S_n, σ_n, η_n). Assume to the contrary that there are strategies $s, s' \in S_n$ such that the former has a higher average probability of defection, i.e., $\bar{\eta}_s(d) > \bar{\eta}_{s'}(d)$. Let $\alpha = \bar{\eta}_s(d) - \bar{\eta}_{s'}(d)$. Note that agents who follow strategy s have a strictly higher payoff than agents who follow s' when being matched with normal partners. This is because strategy s yields: (1) a strictly higher direct payoff of at least $\alpha \cdot l$ due to playing more often the dominant action d, and (2) a weakly higher payoff against normal agents, because the fact that it defects more often and all normal agents follow defector-favoring strategies implies that normal partners defect with a weakly smaller probability

when being matched with agents who follow strategy s (relative to s'). We also need to consider what happens when normal agents are matched with committed agents. The maximal indirect gain that followers of strategy s' have relative to followers of strategy s, due to inducing a higher probability of cooperation from committed partners, is at most $\epsilon_n \cdot (l+1) \cdot k \cdot \alpha$. This implies that if $\epsilon_n < \frac{l}{(l+1)\cdot k}$, then followers of strategy s have a strictly higher payoff than followers of s', which contradicts (S_n, σ_n, η_n) being a Nash equilibrium.

Third, we argue that for any normal agent it is the case that the probability that the partner defects conditional on the agent observing message $m = k$ is weakly larger than the probability that the partner defects conditional on the agent observing any message $m < k$. To see why this is the case, note that the regularity of the set of commitments implies that not all commitment 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 normal agents defect with the same probability, which implies that they induce the same signal distribution, and thus they induce the same behavior from all partners. Combining this fact with the fact that not all commitment strategies have the same defection probability, implies (for a sufficiently small ϵ_n) that if a player observes a message that includes only defections, then the partner is more likely to have a higher average defection probability against normal agents (i.e., $q(m, s) < q(k, s)$ for any normal strategy s and any $m < k$).

Thus, any normal agent (who follows a defector-favoring strategy due to the first step) defects with a weakly higher probability after observing signal $m = k$. This implies that if ϵ_n is sufficiently small, then a deviator who always defects outperforms the incumbents. The deviator achieves a direct higher payoff by defecting more often, as well as a weakly higher indirect gain by inducing the incumbents to cooperate more often.

C.7 Proof of Theorem 2 (Stable Cooperation in Defensive PDs)

Part 1: Let $(S^*, \sigma^*, \eta^* \equiv c)$ be a perfect equilibrium. This implies that there exist a distribution of commitments (S_C, λ), a converging sequence of strictly positive commitment levels $\epsilon_n \to_{n\to\infty} 0$, and a converging sequence of steady states $(S_n, \sigma_n, \eta_n) \to_{n\to\infty} (S^*, \sigma^*, \eta^*)$, such that for each n the state (S_n, σ_n, η_n) is a Nash equilibrium of the perturbed environment $((G, k), (S_C, \lambda), \epsilon_n)$. The fact that the equilibrium induces full cooperation (in the limit when $\epsilon_n \to_{n\to\infty} 0$) implies that all normal agents must cooperate when they observe no defections, i.e., $s_0(c) = 1$ for each $s \in S^*$.

Next we show that $s_1(d) > 0$ for some $s \in S^*$. Assume to the contrary that $s_1(d) = 0$ for every $s \in S^*$. This implies that for any $\delta > 0$, if n is sufficiently large then $\sum_{s\in S_n} \sigma_n(s) \cdot s_1(d) < \delta$. It follows that if a deviator (Alice) who follows a strategy s' defects with a small probability of $\alpha << 1$ when observing no defections (i.e., $s'_0(d) = \alpha$), then she outperforms the incumbents. To see this note that since she occasionally defects when observing $m = 0$ she obtains a direct gain of at least $\alpha \cdot g \cdot \Pr(m = 0)$, where $\Pr(m = 0)$ is the probability of observing $m = 0$ given the steady state (S_n, σ_n, η_n). The probability that a partner observes her defecting twice or more is $O(\alpha^2)$. This implies that her indirect loss from these defections is at most $\left(O(\alpha^2) + O(\alpha) \cdot O(\delta + \epsilon_n)\right) \cdot (1 + l)$ and, thus, for sufficiently small values $\alpha, \delta, \epsilon_n > 0$, Alice strictly outperforms the incumbents.

We now show that $s_m(d) = 0$ for all $s \in S^*$ and all $m \geq 2$. The fact that $\eta^* \equiv c$ implies that for a sufficiently large n, all normal agents cooperate with an average probability very close to one and, thus the average probability of defection by an agent who follows a strategy $s \in S \cup S_C$ is very close to $s_0(d)$. Hence the distribution of signals induced by such an agent is very close to $\nu_{s_0(d)}$. Recall that we assume that the distribution of commitments contains at least one strategy s with $s_0(d) > 0$. This implies that the posterior probability that the partner is going to defect is strictly increasing in the signal m that the agent observes about

the partner. Note that the direct gain from defecting is strictly increasing in the probability that the partner defects as well (due to the game being defensive), while the indirect influence of defection (on the behavior of future partners who may observe the current defection) is independent of the partner's play. From the previous paragraph we know that defection is a best reply conditional on an agent observing $m = 1$. This implies that defection must be the unique best reply when an agent observes at least two defections (i.e., when $m \geq 2$).

It remains to show that there is a normal incumbent strategy that cooperates with positive probability after observing a single defection, i.e., $s_1(d) < 1$ for some $s \in S^*$. Assume to the contrary that $s_1(d) = 1$ for every $s \in S^*$. Let r_n denote the average probability that a normal agent defects after observing $m \geq 1$. Since $(S_n, \sigma_n, \eta_n) \to_{n\to\infty} (S^*, \sigma^*, \eta^*)$, the assumption that $s_1(d) = 1$ for all $s \in S^*$ implies that $r_n > 0.6$ for a sufficiently large n. Let $Pr(m \geq 1|S_n)$ denote the probability of observing $m \geq 1$ conditional on being matched with a normal partner. Note that the assumption that $\hat{s}_0(d) > 0$ for some committed strategy $\hat{s}$ and the assumption that $s_1(d) > 0$ for some normal strategy together imply that $Pr(m \geq 1|S_n) > 0$. Note that $\eta^* \equiv c$ implies that $\lim_{n\to\infty}(Pr(m = 1|S_n)) = 0$. Hence $Pr(m = 1|S_n)$ is $O(\epsilon_n)$. We can calculate $Pr(m \geq 1|S_n)$ as follows:

$$Pr(m \geq 1|S_n) = k\cdot((1-\epsilon_n)\cdot r_n \cdot Pr(m \geq 1|S_n) + \epsilon_n \cdot \lambda(\hat{s}) \cdot (\hat{s}_0(d) + O(\epsilon_n))) - O\left(\epsilon_n^2\right) - O\left(\left(Pr(m \geq 1|S_n)\right)^2\right).$$

The reason for this equation is as follows. The observed signal induced by a normal agent (Bob) describes his actions in k interactions. In each of these interactions Bob's partner was normal with a probability of $1-\epsilon_n$, and was committed with a probability of ϵ_n. If Bob's partner in an interaction was normal then she defected with a probability of r_n when the partner observed $m \geq 1$ (which happened with a probability of $Pr(m \geq 1|S_n)$). If Bob's partner in an interaction was committed then she followed strategy $\hat{s}$ with a probability of $\lambda(\hat{s})$ and defected with a probability of $\hat{s}_0(d) + O(\epsilon_n)$ (as argued above, the average defection probability of an agent following strategy s should be close to $s_0(d)$). Finally, the terms $-O\left(\epsilon_n^2\right) - O\left(\left(Pr(m \geq 1|S_n)\right)^2\right)$ subtract over-counting cases in which Bob has defected more than once. Rearranging and simplifying the above equation, by using the fact that $\left(Pr(m \geq 1|S_n)\right)^2$ is $O\left(\epsilon_n^2\right)$, yields

$$(1 - k\cdot(1-\epsilon_n)\cdot r_n)\cdot Pr(m \geq 1|S_n) = k\cdot(\epsilon_n \cdot \lambda(\hat{s})\cdot\hat{s}_0(d)).$$

Then use $r_n > 0.6$ to infer that the LHS is negative. This contradicts the fact that the RHS is positive.

Part 2: Recall that s^1 (s^2) is the strategy that induces an agent to defect iff the agent observes $m \geq 1$ ($m \geq 1$). Let s^q be the strategy that induces an agent to defect with a probability of q (to be defined later) iff the agent observes $m = 1$, to defect for sure if she observes $m \geq 2$, and to cooperate for sure if she observes $m = 0$. Let (S_C, λ) be an arbitrary distribution of commitments. We will show that there exist a converging sequence of commitment levels $\epsilon_n \to 0$ and converging sequences of steady states

$$\psi_n \equiv \left(\left\{s^1, s^2\right\}, (q_n, 1-q_n), \eta_n\right) \to_{n\to\infty} \psi^* \equiv \left(\left\{s^1, s^2\right\}, (q, 1-q), \eta \equiv c\right),$$

and

$$\psi'_n \equiv \left(\left\{s^{q_n}\right\}, 1_{s^{q_n}}, \eta'_n\right) \to_{n\to\infty} \psi'^* \equiv \left(\left\{s^q\right\}, 1_{s^q}, \eta' \equiv c\right),$$

such that either (1) for each n the steady state ψ_n is a Nash equilibrium of $((G,k),(S_C,\lambda),\epsilon_n)$, or (2) for each n the steady state ψ'_n is a Nash equilibrium of $((G,k),(S_C,\lambda),\epsilon_n)$. This means that cooperation is a strictly perfect equilibrium action.

Fix an $n \geq 1$ such that ϵ_n is sufficiently small. (Exactly what counts as sufficiently small will become clear below.) In what follows, we calculate a number of probabilities while relying on the fact that $\epsilon_n << 1$. Thus we neglect terms of $O(\epsilon_n)$ (resp., $O(\epsilon_n^2)$) when the leading term is $O(1)$ (resp., $O(\epsilon_n)$). The calculations give the same results for ψ_n as for ψ_n'. Since we are looking for consistent behaviors η_n and η_n' such that $\eta_n \to_{n\to\infty} \eta \equiv c$ and $\eta_n' \to_{n\to\infty} \eta' \equiv c$, we assume that $(\eta_n)_{s_i}(s_j)(c) = 1 - O(\epsilon_n)$ for each $s_i, s_j \in \{s^1, s^2\}$ in ψ_n and assume that $(\eta_n')_{s^q}(s^q)(c) = 1 - O(\epsilon_n)$ in ψ_n'. We later confirm that there indeed exist consistent behaviors η_n and η_n' with these properties.

For each incumbent strategy s, let $Pr(m = 1|s)$ ($Pr(m \geq 2|s)$) denote the probability of observing exactly one defection (at least two defections) conditional on the partner following strategy s. Let $Pr(m = 1)$ and $Pr(m \geq 2)$ be the corresponding unconditional probabilities.

The assumption that $\eta_n \to_{n\to\infty} \eta \equiv c$ and $\eta_n' \to_{n\to\infty} \eta' \equiv c$ implies that agents are very likely to observe the message $m = 0$ (i.e., zero defections) when being matched with a random partner. Formally:

$$Pr(m = 0) = (1 - O(\epsilon_n))^k = 1 - O(\epsilon_n).$$

The conditional probabilities of observing $m = 0$, $m = 1$, and $m \geq 2$, for all $s \in S_n \cup S_C$, are

$$Pr(m = 0|s) = (s_0(c))^k + O(\epsilon_n),$$

$$Pr(m = 1|s) = k \cdot s_0(d) \cdot (s_0(c))^{k-1} + O(\epsilon_n),$$

$$Pr(m \geq 2|s) = 1 - Pr(m = 0|s) - Pr(m = 1|s).$$

Let $S_n = \{s^1, s^2\}$ in ψ_n and $S_n = \{s^{q_n}\}$ in ψ_n'. Given message m, let $Pr(m|S_n)$ denote the probability of observing message m, conditional on the partner following a normal strategy. Specifically, in the heterogeneous state ψ_n (with two normal strategies), this conditional probability is given by

$$Pr(m|S_n) = q \cdot Pr(m|s^1) + (1 - q) \cdot Pr(m|s^2).$$

Furthermore, it follows (from the expressions for $Pr(m = 0|s)$, $Pr(m = 1|s)$, and $Pr(m \geq 2|s)$ that

$$Pr(m = 0|S_n) = 1 - O(\epsilon_n), \quad Pr(m = 1|S_n) = O(\epsilon_n) \quad Pr(m \geq 2|S_n) = O(\epsilon_n^2).$$

Next we calculate the probability that a normal agent (Alice) generates a message that contains a single defection. This happens with probability one if exactly one of the k interactions sampled from Alice's past was such that Alice observed her partner in that interaction to have defected at least twice (which implies that her partner is most likely to have been a committed agent). This happens with probability q_n if exactly one of the k interactions sampled from Alice's past was such that Alice observed her partner to have defected exactly once (in which her partner might have been either a committed or a normal agent):

$$\begin{aligned} Pr(m = 1|S_n) &= k \cdot \sum_{s \in S_C} \epsilon_n \cdot \lambda(s) \cdot (Pr(m \geq 2|s) + q_n \cdot Pr(m = 1|s)). \\ &\quad + k \cdot (1 - \epsilon_n) \cdot [q_n \cdot Pr(m = 1|S_n) + Pr(m \geq 2|S_n)] \\ &\quad + O(\epsilon_n^2). \end{aligned}$$

The final term $O(\epsilon_n^2)$ comes from the very small probability of observing a normal agent to defect twice. Since

$Pr(m=1|S_n) = O(\epsilon_n)$ and $Pr(m \geq 2|S_n) = O(\epsilon_n^2)$, this can be simplified (neglecting $O(\epsilon_n^2)$) and rearranged to obtain

$$Pr(m=1|S_n) = \frac{k \cdot \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot (Pr(m \geq 2|s) + q_n \cdot Pr(m=1|s))}{1 - k \cdot q}, \tag{17}$$

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

$$Pr(m=1) = \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s) + Pr(m=1|S_n) + O(\epsilon_n^2),$$

$$\begin{aligned} Pr(m \geq 2) &= \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot Pr(m \geq 2|s) + (1-\epsilon_n) \cdot Pr(m \geq 2|S_n) \\ &= \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot Pr(m \geq 2|s) + O(\epsilon_n^2). \end{aligned}$$

By using Bayes' rule we can calculate the conditional probability that the partner uses strategy $s \in S_C$ as a function of the observed message:

$$\begin{aligned} Pr(s|m=0) &= \frac{\epsilon_n \cdot \lambda(s) \cdot Pr(m=0|s)}{Pr(m=0)}, \\ Pr(s|m=1) &= \frac{\epsilon_n \cdot \lambda(s) \cdot Pr(m=1|s)}{Pr(m=1)}, \\ Pr(s|m \geq 2) &= \frac{\epsilon_n \cdot \lambda(s) \cdot Pr(m \geq 2|s)}{Pr(m \geq 2)}. \end{aligned}$$

Note that

$$\sum_{s \in S_C} Pr(s|m=0) = \frac{\epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot (s_0(c))^k}{1 - O(\epsilon_n)} = O(\epsilon_n).$$

From Eq. (17) we have

$$\sum_{s \in S_n} \sigma(s) \cdot Pr(m=1|s) = Pr(m=1|S_n) = \frac{k \cdot \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot (Pr(m \geq 2|s) + q \cdot Pr(m=1|s))}{1 - k \cdot q_n}.$$

We use this to obtain, by Bayes' rule,

$$\begin{aligned} \sum_{s \in S_C} Pr(s|m=1) &= \frac{\epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s)}{\epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s) + \frac{k \cdot \epsilon_n \cdot \sum_{s \in S_C} \lambda(s) \cdot (Pr(m \geq 2|s) + q_n \cdot Pr(m=1|s))}{1 - k \cdot q_n} + O(\epsilon_n^2)} \\ &= \frac{\sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s)}{\sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s) + \frac{k \cdot \sum_{s \in S_C} \lambda(s) \cdot (Pr(m \geq 2|s) + q_n \cdot Pr(m=1|s))}{1 - k \cdot q_n} + O(\epsilon_n^2)} \end{aligned}$$

Note that the terms $\sum_{s \in S_C} \lambda(s) \cdot Pr(m=1|s)$ and $\sum_{s \in S_C} \lambda(s) \cdot (Pr(m \geq 2|s))$ do not vanish as $\epsilon_n \to 0$. Moreover we will see below (Eqs. (19) and (20)) that this implies that q_n also not vanish as $\epsilon_n \to 0$. Together these observations imply that there are numbers $a, b \in (0,1)$ such that, for all n, it is the case that

$$0 < a < \sum_{s \in S_C} Pr(s|m=1) < b < 1. \tag{18}$$

Furthermore

$$\sum_{s\in S_C} Pr\left(s|m\geq 2\right)=\frac{1}{1+\frac{\sum_{s\in S_n}\sigma(s)\cdot Pr(m\geq 2|s)}{\epsilon_n\cdot\sum_{s\in S_C}\lambda(s)\cdot Pr(m\geq 2|s)}}=\frac{1}{1+\frac{O(\epsilon_n^2)}{O(\epsilon_n)}}=\frac{1}{1+O\left(\epsilon_n\right)}.$$

Hence for a sufficiently large n, the more defections there are in the observed message, the higher is the conditional probability that the partner is committed:

$$\sum_{s\in S_C} Pr\left(s|m=0\right)<\sum_{s\in S_C} Pr\left(s|m=1\right)<\sum_{s\in S_C} Pr\left(s|m\geq 2\right).$$

Let $Pr\left(S_n|m=1\right)=\sum_{s\in S_n} Pr\left(s|m=1\right)$ denote the conditional probability that the partner follows a normal strategy conditional on the agent observing message $m=1$. Eq. (18) implies that there are numbers $a',b'\in(0,1)$ such that, for all n, it is the case that $0<a'<Pr\left(S_n|m=1\right)<b'<1$ (because $Pr\left(S_n|m=1\right)+\sum_{s\in S_C} Pr\left(s|m=1\right)=1$).

Let μ_n be the probability that a random partner defects conditional on a player observing message $m=1$ about the partner, and conditional on the partner observing the message $m=0$:

$$\mu_n=\sum_{s\in S_C} Pr\left(s|m=1\right)\cdot s_0\left(d\right)+O\left(\epsilon_n\right). \tag{19}$$

Eq. (19) defines μ_n as a strictly decreasing function of q_n. To see this, note that the term $s_0\left(d\right)$ does not depend on q_n, and in $Pr\left(s|m=1\right)=\frac{\epsilon_n\cdot\lambda(s)\cdot Pr(m=1|s)}{Pr(m=1)}$ the numerator does not depend on q_n, whereas the term $Pr\left(m=1\right)$ is increasing in q_n.

Next we calculate the value of q_n that balances the payoff of both actions after a player observes a single defection (neglecting terms of $O\left(\epsilon_n^2\right)$). The LHS of the following equation represents the player's direct gain from defecting when she 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:

$$\Pr\left(m=1\right)\cdot\left(\mu_n\cdot l+\left(1-\mu_n\right)\cdot g\right)=\Pr\left(m=1\right)\cdot\left(k\cdot q\cdot\left(l+1\right)+O\left(\epsilon_n\right)\right)\Rightarrow q_n=\frac{\mu_n\cdot l+\left(1-\mu_n\right)\cdot g}{k\cdot\left(l+1\right)}+O\left(\epsilon_n\right). \tag{20}$$

Note that Eq. (20) defines q_n as a strictly increasing function of μ_n. This implies that there are unique values of q_n and μ_n satisfying $\frac{g}{k\cdot(l+1)}<q_n<\frac{l}{k\cdot(l+1)}<\frac{1}{k}$ and $0<\mu_n<1$, which jointly solve Eqs. (19) and (20). This pair of parameters balances the payoff of both actions when a player observes a signal $m=1$. Note that sequences of $\left(q_n\right)_n\to q$ and $\left(\mu_n\right)_n\to\mu$ converge to the values that solve the above equations when ignoring the terms that are $O\left(\epsilon_n\right)$.

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. (20), and the indirect loss is still given by the RHS of Eq. (20). The reason that the direct gain is larger is that normal partners almost never defect twice or more (the probability is $O\left(\epsilon_n^2\right)$), and thus the partner is most likely be committed and to defect with a probability that is higher than μ_n (since μ_n also gives weight to normal strategies that are most likely to cooperate). More generally, note that given that the normal agents almost always cooperate, the average probability of defection of each agent who follows strategy s is $s_0\left(d\right)+O\left(\epsilon_n\right)$. This implies that for a sufficient small ϵ_n, the higher m is, the higher the partner's value of $s_0\left(d\right)$ is likely to be. Hence the higher m is, the higher is the probability that the partner will defect against a normal agent. Thus the direct gain from defection is increasing in the signal m

that the normal agent observes about her partner. (A formal detailed proof of this statement is available upon request.)

Next, consider a deviator (Alice) who defects with a probability of $\alpha > 0$ after she observes $m = 0$. In what follows we calculate Alice's expected payoff as a function of α in any post-deviation stable state, neglecting terms of $O(\epsilon_n)$ throughout the calculation. Note that Alice's partner observes signal $m = 1$ with a probability of $k \cdot \alpha \cdot (1-\alpha)^{k-1}$, and observes signal $m \geq 2$ with a probability of $1 - (1-\alpha)^k - k \cdot \alpha \cdot (1-\alpha)^{k-1}$. This implies that the mean probability that a normal partner defects against a 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 $\frac{\partial \pi(\alpha)}{\partial \alpha}$ reveals that $\pi(\alpha)$ is strictly decreasing in α for each $q > \frac{g}{k \cdot (l+1)}$. Thus any deviator with $\alpha > 0$ earns strictly less than the incumbents (who have $\alpha = 0$).

We have now shown that the best reply is c after observing $m = 0$ and d after observing $m \geq 2$. After observing $m = 1$ both c and d are best replies provided that q has the required value. That is, we know what the aggregate probability of defection after observing $m = 1$ has to be in equilibrium. However, we do not know whether mixing will occur at the individual level. We now turn to this question.

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 $\epsilon_n \to 0$):

$$\chi = \lim_{n \to \infty} \left(\sum_{s \in S_C} Pr(s|m=1) \cdot s^1(d) + Pr(S_n|m=1) \cdot q \right).$$

We conclude by showing that if $\chi > \mu$ ($\chi < \mu$), then ψ^* (ψ'^*) is a perfect equilibrium. This is so because if $\chi > \mu$ ($\chi < \mu$), then conditional on a normal agent observing a single defection, the partner is more (less) likely to defect the higher the probability with which the agent defects when she 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, the higher the agent's own defection probability is, the more profitable defection is (recall that the higher the probability of defection of the partner is, the higher the direct gain from defection, whereas 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 a deviator who mixes on the individual level (defects with probabilities different from q) is outperformed when $\chi > \mu$ ($\chi < \mu$)).

Note that the normal agents are more likely to defect against a partner who is more likely to defect when she observes a single defection. This implies that when focusing only on normal partners, the induced level of χ is larger than the induced level of μ. It is only the committed agents who may induce the opposite inequality (namely, $\chi < \mu$). Thus, if in the limit as $\epsilon \to 0$ the equality $\chi = \mu$ holds, then it must be that for any positive small share of committed agents ϵ_n, it is the case that $\chi_n < \mu_n$, which implies by the argument above that the state ψ'_n is a Nash equilibrium.

Remark 8. The above argument shows that when $\chi < \mu$, each state ψ'_n is a *strictly* perfect equilibrium (any

deviator who follows a strategy different from s^{q_n} obtains a strictly lower payoff). In the opposite case of $\chi > \mu$ one can show that an agent who follows strategy s_i achieves a higher payoff than an agent who follows s_{-i} conditional on the partner following s_i. This implies that the mixed equilibrium between the strategies of s^1 and x_2 is Hawk-Dove-like, and that the state ψ_n is evolutionarily stable (see Appendix B). This shows that cooperation is robust also to joint deviation of a small group of agents, and that it satisfies the refinement of evolutionary stability defined in Appendix B (namely, cooperation is a strictly perfect evolutionarily stable action).

C.8 Proof of Proposition 5 (Observing a Single Action)

Arguments and pieces of notation that are analogous to the ones used in the proof of Theorem 2 are presented in brief or skipped. Let $s^c \equiv c$ be the strategy that always cooperates. The same arguments as in Theorem 2 show that the only possible candidates to be perfect equilibria that support full cooperation, are steady states of the form $\psi = \left(\left\{s^1, s^c\right\}, (q, 1-q), \eta \equiv c\right)$ or $\psi' = \left(\left\{s^q\right\}, 1_{s^q}, \eta' \equiv c\right)$.

Consider a perturbed environment $\left((G_{PD}, k), (S_C, \lambda), \epsilon\right)$ where $\epsilon > 0$ is sufficiently small. In what follows: (1) for the case of $g \leq \beta_{C,\lambda}$ we characterize a Nash equilibrium of this perturbed environment that is within a distance of $O(\epsilon)$ from either ψ or ψ', and (2) we show that no such Nash equilibrium exists for the case of $g > \beta_{C,\lambda}$.

Consider a steady state that is within a distance of $O(\epsilon)$ from either ψ or ψ'. The fact that the behavior in the steady state is close to always cooperating (i.e., to $\eta \equiv c$) implies that the probability of observing $m = 1$ conditional on the partner following a commitment strategy $s \in S_C$ is:

$$Pr(m = 1|s) = s_0(d) + O(\epsilon).$$

Similarly, the probability of observing $m = 1$ conditional on the partner being normal is

$$Pr(m = 1|S_n) = q \cdot \left(\epsilon \cdot \sum_{s \in S_C} \lambda(s) \cdot s_0(d) + (1-\epsilon) \cdot Pr(m = 1|S_n)\right) + O\left(\epsilon^2\right) \Rightarrow$$

$$Pr(m = 1|S_n) = \frac{\epsilon \cdot q \cdot \sum_{s \in S_C} \lambda(s) \cdot s_0(d)}{1-q} + O\left(\epsilon^2\right).$$

By using Bayes' rule we can calculate the probability that the partner uses strategy $s \in S_C$ conditional on observing $m = 1$:

$$Pr(s|m = 1) = \frac{\epsilon_n \cdot \lambda(s) \cdot Pr(m = 1|s)}{Pr(m = 1)} = \frac{\epsilon \cdot \lambda(s) \cdot s_0(d)}{\epsilon \cdot \left(\sum_{s \in S_C} \lambda(s) \cdot s_0(d) + \frac{q \cdot \sum_{s \in S_C} \lambda(s) \cdot s_0(d)}{1-q}\right)} + O(\epsilon) \Rightarrow$$

$$Pr(s|m = 1) = \frac{(1-q) \cdot \lambda(s) \cdot s_0(d)}{\sum_{s \in S_C} \lambda(s) \cdot s_0(d)} + O(\epsilon).$$

Let μ be the probability that a random partner defects conditional on an agent observing message $m = 1$ about the partner, and conditional on the partner observing the message $m = 0$ about the agent. (Note that only committed partners defect with positive probability when observing $m = 0$.)

$$\mu = \sum_{s \in S_C} Pr(s|m = 1) \cdot s_0(d) + O(\epsilon) = (1-q) \cdot \frac{\sum_{s \in S_C} \lambda(s) \cdot (s_0(d))^2}{\sum_{s \in S_C} \lambda(s) \cdot s_0(d)} + O(\epsilon) = (1-q) \cdot \beta_{(S_C,\lambda)} + O(\epsilon). \quad (21)$$

Next we calculate the value of q that balances the payoff of both actions after a player observes a single defection. The LHS of the following equation represents the player's direct gain from defecting when she observes a single defection, while the RHS represents the player's indirect loss induced by future partners who defect as a result of observing these defections:

$$\Pr(m=1)\cdot(\mu\cdot l+(1-\mu)\cdot g)\quad +O(\epsilon)=\quad \Pr(m=1)\cdot(q\cdot(l+1)+O(\epsilon))\Rightarrow. \tag{22}$$

$$q=\frac{\mu\cdot l+(1-\mu)\cdot g}{l+1}+O(\epsilon)=\frac{g+\mu\cdot(l-g)}{l+1}+O(\epsilon) \tag{23}$$

Substituting (21) in (23) yields:

$$q=\frac{g+(1-q)\cdot(l-g)\cdot\beta_{(S_C,\lambda)}}{l+1}+O(\epsilon)\Rightarrow q\cdot(l+1)=g+(1-q)\cdot(l-g)\cdot\beta_{(S_C,\lambda)}+O(\epsilon)$$

$$\Rightarrow q=\frac{g+(l-g)\cdot\beta_{(S_C,\lambda)}}{l+1+(l-g)\cdot\beta_{(S_C,\lambda)}}+O(\epsilon).$$

Consider a deviator (Alice) who always defects. Normal partners of Alice cooperate with probability of $1-q$. This implies that Alice gets an expected payoff of $(1+g)\cdot(1-q)$, while the normal agents each get a payoff of $1+O(\epsilon)$. Alice is outperformed iff (neglecting terms of $O(\epsilon)$):

$$(1+g)\cdot(1-q)\leq 1\Leftrightarrow q\geq\frac{g}{1+g}\Leftrightarrow\frac{g+(l-g)\cdot\beta_{(S_C,\lambda)}}{l+1+(l-g)\cdot\beta_{(S_C,\lambda)}}\geq\frac{g}{1+g}$$

$$\Leftrightarrow(1+g)\cdot\left(g+(l-g)\cdot\beta_{(S_C,\lambda)}\right)\geq g\cdot\left(l+1+(l-g)\cdot\beta_{(S_C,\lambda)}\right)$$

$$\Leftrightarrow g^2+(l-g)\cdot\beta_{(S_C,\lambda)}+\geq g\cdot l\ \Leftrightarrow\ g\cdot(l-g)\leq(l-g)\cdot\beta_{(S_C,\lambda)}\ \Leftrightarrow\ g\leq\beta_{(S_C,\lambda)}.$$

Thus, the steady state can be a Nash equilibrium only if $g\leq\beta_{(S_C,\lambda)}$. It is relatively straightforward to show that if $g\leq\beta_{(S_C,\lambda)}$, then a deviator who defects with probability α when observing $m=0$ is outperformed. The remaining steps of the proof are as in the proof of part (2) of Theorem 2, and omitted for brevity.

C.9 Proof of Theorem 3 (Observing Conflicts)

The proof of part (1a) is analogous to Theorem 2 and is omitted for brevity. We now prove Part 1(b), i.e., that any mild game admits a strictly perfectly equilibrium action. Arguments and notations that are analogous to the proof of Theorem 2 are presented in brief. Let s^1 (s^2) be the strategy that instructs a player to defect if and only if she receives a message containing one or more (two or more) conflicts. Consider the following candidate for a perfect equilibrium $\left(\left\{s^1,s^2\right\},(q,1-q),c\right)$. Here, the probability q will be determined such that both actions are best replies when observing a single conflict.

Let $(\mathcal{S}_\mathcal{C},\lambda)$ be a distribution of commitments. We show that there exists a converging sequence of levels $\epsilon_n\to 0$, and converging sequences of steady states $\left(\left\{s^1,s^2\right\},(q_n,1-q_n),\eta_n\right)\to\left(\left\{s^1,s^2\right\},(q,1-q),\eta\equiv c\right)$ and $\left(\{s^{q_n}\},1_{s^{q_n}},\eta_n'\right)\to\left(\{s^q\},1_{s^q},\eta'\equiv c\right)$ such that either (1) each steady state $\psi_n\equiv\left(\left\{s^1,s^2\right\},\sigma_n\equiv(q_n,1-q_n),\eta_n\right)$ is a Nash equilibrium of $((G,k),(S_C,\lambda),\epsilon_n)$, or (2) each steady state $\psi_n'\equiv\left(\{s^{q_n}\},\sigma_n'\equiv 1_{s^{q_n}},\eta_n'\right)$ is a Nash equilibrium of $((G,k),(S_C,\lambda),\epsilon_n)$.

Fix $n\geq 1$. Assume that ϵ_n is sufficiently small. We calculate the probability $Pr(m=1|S_n)$ that a normal agent (Alice) induces a message $m=1$. Since we focus on the steady states in which the incumbents defect very rarely (i.e., η_n and η_n' converge to $\eta^*\equiv c$), we can assume that $Pr(m=1|S_n)$ is $O(\epsilon_n)$. Alice may be

involved in a conflict if one of her k partners is committed, which happens with a probability of $O(\epsilon_n)$. If all of the k partners are normal, then at each interaction both Alice and her partner defects with a probability of $Pr(m=1|S_n)$, which implies that the probability of a conflict is $2 \cdot Pr(m=1|S_n) - \left(Pr(m=1|S_n)^2\right)$. Therefore:

$$Pr(m=1|S_n) = k \cdot \left(O(\epsilon_n) + 2 \cdot q_n \cdot Pr(m=1|S_n) - O\left(\left(Pr(m=1|S_n)\right)^2\right)\right).$$

Solving this yields, while neglecting terms that are $O(\epsilon_n^2)$ (including $Pr(m=1|S_n)^2$), yields:

$$Pr(m=1|S_n) = \frac{k \cdot O(\epsilon_n)}{1 - 2 \cdot k \cdot q_n}, \tag{24}$$

which is well defined and $O(\epsilon_n)$ as long as $q_n < \frac{1}{2 \cdot k}$. Note that as q_n approaches $\frac{1}{2 \cdot k}$, the value of $Pr(m=1|S_n)$ "explodes" (become arbitrarily larger than terms that are $O(\epsilon_n)$).

By Bayes' rule we can calculate the conditional probability $Pr(s|m=1)$ of being matched with each strategy $s \in S_C$ (same calculations as detailed in the proof of Theorem 2). Note that these conditional probabilities are decreasing in $Pr(m=1|S_n)$, and thus decreasing in q_n. Let μ_n be the probability that a random partner defects conditional on a player observing message $m=1$ about the partner, and conditional on the partner observing the message $m=0$:

$$\mu_n = \sum_{s \in S_C} Pr(s|m=1) \cdot s_0(d) + O(\epsilon_n). \tag{25}$$

Note that μ_n is decreasing in q_n. Moreover, as $q_n \nearrow \frac{1}{2 \cdot k}$, we have $\mu_n(q_n) \searrow 0$, because $Pr(m=1|S_n)$ "explodes" as we approach the threshold of $k \cdot q = 0.5$.

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

$$\Pr(m=1) \cdot (\mu_n \cdot l + (1-\mu_n) \cdot g) = \Pr(m=1) \cdot (1-\mu_n) \cdot k \cdot q \cdot (l+1) + O(\epsilon_n) \quad \Leftrightarrow \quad q_n = \frac{\mu_n \cdot l + (1-\mu_n) \cdot g}{(1-\mu) \cdot k \cdot (l+1)} + O(\epsilon_n). \tag{26}$$

In connection with Eq. (26) it was noted that $q(\mu)$ is increasing in μ_n, and since the game is mild we have $q_n(0) = \frac{g}{k \cdot (l+1)} < \frac{1}{2 \cdot k}$. This implies that there is a unique pair of values of $q_n \in \left(\frac{g}{k \cdot (l+1)}, \frac{1}{2 \cdot k}\right)$ and $\mu_n \in (0,1)$ that jointly solve Eqs. (25) and (26). This pair of values balances the payoff of both actions when a player observes a signal $m=1$. Note that sequences of $(q_n)_n \to q$ and $(\mu_n)_n \to \mu$ converge to the values that solve the above equations when ignoring the terms that are $O(\epsilon_n)$. The remaining arguments of Part (1) are analogous to those in the final part of the proof of Theorem 2, and are omitted for brevity.

Next, we deal with Part (2), namely, the case of an acute Prisoner's Dilemma ($g > 0.5 \cdot (l+1)$). Assume (in order to obtain a contradiction) that the environment admits a perfect equilibrium $(S^*, \sigma^*, \eta^* \equiv c)$. That is, there exists a converging sequence of strictly positive commitment levels $\epsilon_n \to_{n \to \infty} 0$, and a converging sequence of steady states $(S_n, \sigma_n, \eta_n) \to_{n \to \infty} (S^*, \sigma^*, \eta^*)$, such that each state (S_n, σ_n, η_n) is a Nash equilibrium of the perturbed environment $((G,k),(S_C,\lambda),\epsilon_n)$. By the arguments of part (1) (and the arguments of part (1a) of Theorem 2), the average probability q_n by which a normal agent defects when observing $m=1$ in the steady

state (S_n, σ_n, η_n) (for a sufficiently small ϵ_n) should be at least equal to the minimal solution of Eq. (26): $q_n(\mu_n = 0) = \frac{g}{k\cdot(l+1)} + O(\epsilon_n)$. However, if the game is acute, then this minimal solution is larger than $\frac{1}{2\cdot k}$, and Eq. (24) cannot be satisfied by $Pr(m = 1|S_n) << 1$, which yields a contradiction.

C.10 Proof of Theorem 4 (Observing Action Profiles)

Recall, that a signal $m \in M$ consists of information about the number of times in which each of the possible four action profiles have been played in the sampled k interactions. Let $u(m)$ be the number of sampled interactions in which the partner has been the sole defector, and let $d(m)$ denote the number of sampled interactions in which at least of one of the players have defected. Let s^1 and s^2 be defined as follows:

$$s^1(m) = \begin{cases} d & u(m) = 1 \text{ or } d(m) \geq 2 \\ c & \text{otherwise} \end{cases} \quad s^2(m) = \begin{cases} d & d(m) \geq 2 \\ c & \text{otherwise.} \end{cases}$$

That is, both strategies induce agents to defect if the partner has been involved in at least two interactions in which the outcome has not been mutual cooperation. In addition, agents who follow s^1 defect also when observing the partner to be the sole defector in a single interaction.

Assume first that G_{PD} is mild (i.e., $g \leq \frac{l+1}{2}$). Fix a small probability of $0 < \alpha << \frac{1}{k}$. Let $s^\alpha \equiv \alpha$ be the strategy that defects with a probability of α regardless of the signal. In what follows, we show that there exist a converging sequence of commitment levels $\epsilon_n \to 0$ and converging sequences of steady states $\psi_n \equiv \left(\left\{s^1, s^2\right\}, (q_n, 1-q_n), \eta_n\right) \to \left(\left\{s^1, s^2\right\}, (q, 1-q), \eta \equiv c\right)$, such that each steady state ψ_n is a Nash equilibrium of $((G,k), (\{s^\alpha\}, 1_{s^\alpha}), \epsilon_n)$.

Fix a sufficiently small $\epsilon_n << 1$. Let μ_n be the probability that the partner defects conditional on: (1) the agent observing a single unilateral defection and $k-1$ mutual cooperations, i.e., $\hat{m} = \left\{(d,c), \overrightarrow{(c,c)}\right\}$ ($u(m) = d(m) = 1$), and (2) the partner observing k mutual cooperations. The parameter q_n is defined such that it balances the direct gain of defection (LHS of the equation) and its indirect loss (RHS) for a normal agent who almost always cooperates:

$$\Pr(\hat{m})\cdot\mu_n\cdot l + (1-\mu)\cdot g = \Pr(\hat{m})\cdot(1-\mu_n)\cdot k\cdot q_n\cdot(l+1) + O(\epsilon_n) \quad \Leftrightarrow \quad q_n = \frac{\mu_n\cdot l + (1-\mu_n)\cdot g}{(1-\mu_n)\cdot k\cdot(l+1)} + O(\epsilon_n). \quad (27)$$

The equation is the same as in the case of observation of conflicts (see (26) above). In particular, note that the indirect cost of defection when the current partner cooperates is only $O(\epsilon_n)$, because it influences only the behavior of normal future partners if they observe an additional interaction different from (c,c) in the k sampled interactions, which only happens with probability of $O(\epsilon_n)$. Next, note that $\mu_n = \alpha \cdot Pr\left(s^\alpha | (d,c), \overrightarrow{(c,c)}\right) + O(\epsilon_n)$ because the only agents who follow s^α defect with positive probability when observing k mutual cooperations. Substituting this in (27) yields:

$$q_n = \frac{g + \alpha \cdot Pr\left(s^\alpha | (d,c), \overrightarrow{(c,c)}\right)\cdot(l-g)}{\left(1 - \alpha\cdot Pr\left(s^\alpha | (d,c), \overrightarrow{(c,c)}\right)\right)\cdot k\cdot(l+1)} + O(\epsilon_n) = \frac{g}{k\cdot(l+1)} + O(\alpha) + O(\epsilon_n).$$

The mildness of the game ($g < \frac{l+1}{2}$) implies that $k\cdot q_n < 0.5$.

Let p_n be the average probability with which the normal agents defect when being matched with committed agents. When $\alpha << \frac{1}{k}$, the s^2-agents rarely ($O(\alpha^2)$) defect against the committed agents, because it is rare to observe these committed agents defecting more than once. The s^1-agents defect against the committed agents

with a probability of $k \cdot q_n \cdot \alpha + O\left(\alpha^2\right) + O\left(\epsilon_n\right)$ because each rare defection of the committed agents is observed with a probability of $k \cdot q$ by s^1-agents. Since $\alpha, p_n << 1$, bilateral defections are very rare ($O\left(\alpha^2\right)$). This implies that $p_n = \alpha \cdot k \cdot q_n + O\left(\alpha^2\right) + O\left(\epsilon_n\right) < \frac{\alpha}{2}$.

Let r_n be the probability that a s^1-agent defects against a fellow s^1-agent. In each observed interaction, the s^1 partner interacts with a committed (resp., s^1, s^2) opponent with a probability of ϵ_n (resp., q_n, 1-q_n) and the partner unilaterally defects with a probability of $\alpha \cdot k \cdot q_n + O\left(\epsilon_n\right) + O\left(\alpha^2\right)$ (resp., $r_n + O\left(r_n^2\right)$, $O\left(\epsilon_n \cdot \alpha^2\right)$). This implies that r_n solves the following equation:

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

where the latter inequality is because $k \cdot q_n < 0.5$. The above calculations show that the total frequency with which committed agents unilaterally defect ($\alpha \cdot \epsilon_n$) is higher than the total frequency with which normal agents unilaterally defect ($q_n + p_n \cdot \delta_n < \alpha \cdot \epsilon_n$). This implies that the probability that an agent is committed, conditional on his being the sole defector in an interaction, is higher than 50%, and that it is higher than this probability conditional on her being the sole cooperator. Next, note that mutual defections between a committed agent and an s^1-agent have a frequency of $O\left(\epsilon_n\right)$, while mutual defections between two committed agents (or two normal agents) are very rare ($O\left(\epsilon_n^2\right)$), which implies that the probability that the partner follows a committed strategy conditional on the player observing mutual defection is $50\% + O\left(\epsilon_n\right)$. This implies that

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

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 committed (because normal agents have two outcomes different from mutual cooperation with a probability of only $O\left(\epsilon_n^2\right)$). This implies that the normal agents play the unique best reply after any signal other than $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$, and thus any deviator who behaves differently in these cases will be outperformed.

Let χ_n be the probability that a random partner defects conditional on both the agent and the partner observing signal $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$. The definitions of strategies s^α, s^1, and s^2 immediately imply that $\chi_n > \mu_n$, and analogous arguments to those presented at the end of the proof of Theorem 2 show that deviators who defect with a probability strictly between zero and one after observing $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$ are outperformed (because an agent's payoff is a strictly convex function of the agent's defection probability when observing signal $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$).

Next assume that the G_{PD} is acute. We have to show that cooperation is not a perfect equilibrium action. Assume to the contrary that $(S^*, \sigma^*, \eta^* \equiv c)$ is a perfect equilibrium WRT distribution of commitments $(\mathcal{S}_\mathcal{C}, \lambda)$. Let $\psi_n = (S_n, \sigma_n, \eta_n) \to (S^*, \sigma^*, c)$ be a converging sequence of Nash equilibria in the converging sequence of perturbed environments $\left(\left(G_{PD}, k\right), \left(\mathcal{S}_\mathcal{C}, \lambda\right), \epsilon_n\right)$. Analogous arguments to the proof of part (1a) of Theorem 2 show that any perfect equilibrium that implements full cooperation $(S^*, \sigma^*, \eta^* \equiv c)$ must satisfy: (1) $s_{\overrightarrow{(c,c)}} = c$ for each $s \in S^*$, (2) if $d(m) \geq 2$ then $s_m = d$ for each $s \in S^*$, and (3) there are $s, s' \in S^*$ such that $s_{\left\{(d,c), \overrightarrow{(c,c)}\right\}}(d) > 0$ and $s_{\left\{(d,c), \overrightarrow{(c,c)}\right\}}(d) < 1$.

Let $0 < q_n < 1$ be the average probability according to which a normal agent defects when she observes $\left\{\left(d,c\right), \overrightarrow{(c,c)}\right\}$. By analogous arguments to those presented above (see (27)) q_n is an increasing function of μ_n, and $q_n\left(\mu_n = 0\right) = \frac{g}{k \cdot (l+1)}$. The acuteness of the game implies that $k \cdot q_n > \frac{g}{(l+1)} \geq > \frac{1}{2}$.

Let $s_\beta \in S_C$ be a committed strategy that induces an agent who follows it (called s_β-agent) to defect with

a probability of $\beta > 0$ when he observes $\left(\overrightarrow{(c,c)}\right)$. In what follows, we show that the presence of strategy s_β induces the normal agents to unilaterally defect more often than s_β-agents. Let p_n be the average probability that normal agents defect against s^α-agents in state ψ_n. This probability p_n must solve the following inequality:

$$\begin{aligned} 1-p_n \quad \geq \quad & ((1-\beta)\cdot(1-p_n))^k + k\cdot((1-\beta)\cdot(1-p_n))^{k-1}\cdot(1-(1-\beta)\cdot(1-p_n)) \\ & + (1-q_n)\cdot k\cdot((1-\beta)\cdot(1-p_n))^{k-1}\cdot\beta\cdot(1-p_n) + O(\epsilon_n). \end{aligned} \tag{28}$$

The LHS of (28) is the average probability that normal agents cooperate against s_β-agents (recall that normal agents always defect when they observe at most $k-2$ mutual cooperations). The normal agents cooperate with probability one (resp., at most one, q_n) 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 $((1-\beta)\cdot(1-p_n))^k$ (resp., $k\cdot((1-\beta)\cdot(1-p_n))^{k-1}\cdot(1-(1-\beta)\cdot(1-p_n))$, $k\cdot((1-\alpha)\cdot(1-p_n))^{k-1}\cdot\alpha\cdot(1-p)$).

Direct numerical analysis of Eq. (28) shows that the minimal p_n that solves this inequality (given that $q_n > \frac{1}{2\cdot k}$) is greater than $\frac{\beta}{2-\beta}$ for any $\beta\in(0,1)$. The total frequency of interactions in which the s_β-agents unilaterally defect is $\beta\cdot(1-p_n)\cdot\epsilon_n\cdot\lambda(s_\beta)+O(\epsilon_n^2)$. The total frequency of interactions in which normal agents unilaterally defect against the s_β-agents is $p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda(s_\beta)+O(\epsilon_n^2)$. Eq. (27) shows that these unilateral defections against s_β-agents induce the normal agents to unilaterally defect among themselves with a total frequency of $\frac{p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda(s_\beta)}{1-k\cdot q_n}+O(\epsilon_n^2) > p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda(s_\beta)$. Finally, note that $p_n > \frac{\beta}{2-\beta} \Leftrightarrow 2\cdot p_n\cdot(1-\beta) > \beta\cdot(1-p_n)$ implies that normal agents unilaterally defect (as the indirect result of the presence of the s_β-agents) more often than s_β-agents.

Next, observe that bilateral defections are most likely to occur in interactions between normal and committed agents. This is because the probability that both normal agents defect against each other is only $O(\epsilon_n^2)$. Thus, when a player observes bilateral defection the partner is more likely to be a committed agent than when the player observes a unilateral defection by the partner. This implies that all the normal 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_n be the (average) probability that normal agents defect when they observe $\left((c,d),\overrightarrow{(c,c)}\right)$. If $w_n < 0.5$, then cooperation is the unique best reply for a normal agent who faces a partner who is likely to defect (e.g., when the normal agent observes 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, $w_n \geq 0.5 \Rightarrow k\cdot w_n > 1$. However, in this case, an analogous argument to the one at the end of the proof of Theorem 3 implies that an arbitrarily small group of mutants who defect with small probability will 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.

C.11 Proof of Theorem 5 (Observing Actions against Cooperation)

The construction of the distribution of commitments $(\{s^\alpha\},1_{s^\alpha})$ and the perfect equilibrium $\left(\{s^1,s^2\},(q,1-q),\eta\equiv c\right)$, and most of the arguments are the same as in the proof of Theorem 4, and are omitted for brevity. Fix ϵ_n sufficiently small. By the same arguments as in the proof of Theorem of 3, the value of q_n that balances the payoffs of s^1 and s^2 satisfy $k\cdot q_n < 1$ for any underlying Prisoner's Dilemma.

Recall that p_n, the average probability with which the normal agents defect when being matched with committed agents, satisfies $p_n = \alpha \cdot k \cdot q_n + O\left(\alpha^2\right) + O\left(\epsilon_n\right) < \alpha$. This implies that the probability that an agent is committed, conditional on her being the sole defector in an interaction, is higher than 50%, conditional on her being the sole cooperator. Next, observe that $\alpha << 1$ implies that the probability $Pr\left((d,d)\right) = O\left(p_n \cdot \alpha^2\right) \cdot O\left(\epsilon_n\right) << Pr\left((c,d)\right) = O\left(p_n \cdot \epsilon_n \cdot \alpha\right)$, which implies that conditional on an agent observing the signal $\left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}$, it is most likely that the partner has cooperated rather than defected in the interaction in which $(*,d)$ has been observed. This implies that $\Pr\left(s^\alpha | \left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}\right) < \Pr\left(s^\alpha | \left\{(d,c), \left(\overrightarrow{(c,c)}\right)\right\}\right)$, and given the value of q_n for which both actions are best replies conditional on observing signal $\left\{(d,c), \left(\overrightarrow{(c,c)}\right)\right\}$, cooperation is the unique best reply when observing either $\left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}$ and $\left\{\left(\overrightarrow{(c,c)}\right)\right\}$, while defection is the unique best reply when observing at most $k-2$ mutual cooperations. This implies that $\left(\left\{s^1, s^2\right\}, (q, 1-q), \eta \equiv c\right)$ is a perfect equilibrium (where q is the limit of q_n when ϵ_n converges to zero.