Comput Econ (2015) 45:469–500
DOI 10.1007/s10614-014-9431-1

A Framework for Computational Strategic Analysis: Applications to Iterated Interdependent Security Games

Yevgeniy Vorobeychik · Steven Kimbrough · Howard Kunreuther

Accepted: 8 March 2014 / Published online: 29 March 2014

Abstract Past work on tournaments in iterated prisoner's dilemma and the evolution of cooperation spawned by Axelrod has contributed insights about achieving cooperation in social dilemmas, as well as a framework for strategic analysis in such settings. We present a broader, more extensive framework for strategic analysis in general games, which we illustrate in the context of a particular social dilemma encountered in interdependent security settings. Our framework is fully quantitative and computational, allowing one to measure the quality of strategic alternatives across a series of measures, and as a function of relevant game parameters. Our special focus on performing analysis over a parametric landscape is motivated by public policy considerations, where possible interventions are modeled as affecting particular parameters of the game. Our findings qualify the touted efficacy of the Tit-for-Tat strategy, demonstrate the importance of monitoring, and exhibit a phase transition in cooperative behavior in response to a manipulation of policy-relevant parameters of the game.

Keywords Tournaments · Game theory · Prisoner's dilemma · Interdependent security games

1 Introduction

Contexts of strategic interaction (games) are ubiquitous. They are often momentous in their consequences. Nations compete with each other in economic markets, for influ-

Y. Vorobeychik (✉)
Vanderbilt University, Nashville, TN, USA
e-mail: eug.vorobey@gmail.com

S. Kimbrough · H. Kunreuther
University of Pennsylvania, Philadelphia, PA, USA

ence on global policy, for prestige that builds “soft power,” and on weapons superiority. Firms compete over consumers, suppliers and access to markets. Individuals interact, competing for salary and promotion within a work environment. Everyone competes over scarce resources. In addition, strategic interactions are often fraught with difficulty for their stakeholders, in part because they are complex and in part because they present challenging tradeoffs between cooperation and competition. Success in war, diplomacy, business, individual advancement, and appropriation of scarce resources may be furthered by well-chosen cooperative alliances.

Two problems of fundamental significance arise for those who would study or participate in such strategic interactions (Kimbrough 2012).

1. Strategy selection. What strategy should a player choose? What are characteristics of good strategies? How should a player go about choosing a strategy of play? The strategy selection problem arises for the individual player, who must decide how to play the game as well, of course, for the scholar who seeks to understand strategic interaction.
2. Institutional design. What outcomes can be expected for a given strategic setup? How might the setup be changed in order to achieve preferred policy outcomes? The institutional design problem arises for the policy maker, and all those who would influence policy.

Beginning with Axelrod’s landmark studies on iterated prisoner’s dilemma (IPD) (Axelrod 1980a, b, 1984; Axelrod and Hamilton 1981) and continuing since, game tournaments—in which collections of strategies for particular games are played with each other and the outcomes analyzed—have contributed to our understanding of the strategy selection and institutional design problems. The tournaments have been studied both analytically and with computerized simulation. IPD games have been the predominant subject of study, although other games are increasingly drawing scholarly attention (Skyrms 2010).

In this paper we report on a series of investigations of Interdependent Security (IDS) games. These games model real-world situations that have important policy significance, where outcomes are stochastic. (See §3 for a detailed discussion.) Moreover, IDS games are unusual in having stochastic payoffs as an essential part of their definition. The stochasticity, by virtue of its magnitude, should not be characterized as mere noise. Recent human subject experiments on IDS games have found that the stochasticity significantly complicates the strategy selection problem for individuals. This in turn leads to challenges for policy formulation.

Extending the results of human subject experiments, we undertake an extensive study of IDS games, using computational tournaments. In doing so, we present a broader, more extensive framework for strategic analysis in general games than is normally realized in existing tournament studies. We illustrate this approach in the context of social dilemmas encountered in IDS settings. Our framework is quantitative and computational, allowing one to measure the quality of strategic alternatives across a series of measures, and as a function of relevant game parameters. We focus particularly on performing analysis over a parametric landscape, motivated by public policy considerations, where possible interventions are modeled as affecting particular parameters of the game. To preview the results, our findings (in the IDS context) qualify the oft-touted efficacy of the Tit-for-Tat strategy, demonstrate the importance of monitoring, and exhibit a phase transition in cooperative behavior in response to a manipulation of policy relevant parameters of the game.

We begin, in the next section, with a discussion of related work that either employs game tournaments or examines games with stochastic elements.

2 Related Work

2.1 Tournaments, Stochasticity

There have been scores of computational game tournaments published since Axelrod's seminal works (Axelrod 1980a, b, 1984; Axelrod and Hamilton 1981). We focus on a few here that are most pertinent to our study, emphasizing tournaments with stochasticity of some kind.

Axelrod's original tournaments are well known, so only the briefest of summaries is needed. Two tournaments for IPD were held with an open solicitation for strategies. The first tournament elicited 14 entries. These 14 plus Random (the strategy of randomly cooperating or defecting at each round) were played and their scores were obtained by summing their points in the pairwise encounters (including play with themselves). Tit-for-Tat won. The results were announced and a second tournament was held. The second tournament garnered 62 entries. These 62 plus Random were played and scored as in the first tournament. Tit-for-Tat won again.

While landmark in their implications, the tournaments and subsequent analyses that elevated Tit-For-Tat as the best strategy in IPD and other such social dilemma encounters have a number of shortcomings.[1]

– The analyses were performed for a single payoff structure of Prisoner's Dilemma games, and it is not clear from Axelrod's studies how sensitive the results were to alternative payoff structures. In the laboratory also, only very few payoff configurations are generally tested, making it difficult to understand the impact of particular payoff choices on the results.
– Although tournament competition is appropriate for identifying a candidate best strategy, in reality the weakest strategies featured in any tournament are unlikely to be played. Consequently, a good strategy should also fare well against a select few high quality strategies, not just all strategies participating in a particular strategic pool.
– While measuring the quality of strategies in terms of their payoffs is quite natural and reasonable, it does not account for uncertainty associated with the specific strategies that would be played by opponents in an actual strategic encounter. One often wishes a strategy to be efficacious against a specific pool of possible opponents and to be robust to variations in particular opponents that could possibly be encountered. At a high level, much of the previous analysis of IPD games that followed in the footsteps of Axelrod focuses on building evidence for Tit-For-Tat as the means for evolution of cooperation, rather than offering an objective strategic

[1] Following publication of the tournaments there ensued a flurry of studies pointing to shortcomings in Tit-for-Tat and offering alternatives, e.g., (Nowak and Sigmund 1993).

analysis with insights for public policy [see, for example Bartholdi et al. (1986); Thomas and Feldman (1988), and numerous other follow-up analyses]. We take this latter perspective.

It is useful to distinguish various possible sources or kinds of stochasticity in games.

1. Incomplete information. This arises when a player is uncertain about the payoffs of its counterpart(s), as may occur, for example, in auctions. The standard treatment of such situations takes a Bayesian approach, yielding Bayesian equilibria and perfect Bayes equilibria as solution concepts. We are not concerned in the present study with incomplete information in this sense.
2. Stochastic strategies. Players may have strategies that involve some form of randomization, as in, for example, mixed strategies in standard game theory. Our study, as do many of the related studies cited in this section, encompasses strategies with such a random component.
3. Accidents, implementation errors. These occur when a player intends to implement one strategy, but ends up playing another. This is commonly referred to as "trembling" or "trembling hand" in the game theory literature (Harsanyi and Selten 1988). This source of stochasticity has been studied extensively, but it is not the subject of our study.
4. Stochasticity in payoffs. This occurs when the game payoff values are random variables. Two kinds of stochastic payoffs have been studied in the literature. The first is *noise,* which occurs as small perturbations in the payoff values realized. Typically (e.g., Bendor 1993; Bendor and Kramer 1991; Rogers et al. 2007), noise is realized by adding a random deviate to the deterministic game payoffs, which is independently and identically distributed for all players and periods [but see Bereby-Meyer and Roth (2006)]. The second is what we will call *strategically conditioned stochasticity,* where the distribution of noise depends on the strategies played by the players. It is present in IDS games and is the main variety of game stochasticity encompassed in our study.

Note on terminology: the literature is inconsistent in the terminology actually deployed. Much of the literature uses *noise* where we would use *accident* or *trembling*. We will indicate when this is the case. Noise in our preferred sense constitutes a form of *stochasticity in payoffs.*

Payoff stochasticity of the first kind—errors of implementation, trembling—has been the subject of several investigations, both analytic and computational. An early analytic study (Fudenberg and Maskin 1990) examined evolutionary stability of strategies for IPD in the face of trembling and found that modest amounts of noise of this type (accidents, trembling) tended to favor cooperation. Bendor developed a different analytic approach and found that both accidents and payoff stochasticity could facilitate cooperation in IPD, although the picture is complex (Bendor 1993). More recently, Pelc and Pelc (2009) undertook an analytic study of strategies in IPD. The focus was primarily on non-stochastic versions. The analysis, however, relied on payoffs in the limit, so that, for example, when Tit-for-Tat plays against Always Defect the two strategies are judged to have identical returns. The paper develops a notion of robustness for pairwise tournaments, which, depending as it does on returns in the limit, may be applied to certain kinds of stochasticity.

On the computational side, Wu and Axelrod considered three main strategies for IPD with respect to "noise" (trembles or accidents in our terminology; Wu and Axelrod 1995). Already at the time (1995) there had been a number of computational tournament studies—cited in the paper—of stochasticity in games brought about by accidents. The three strategies (or families of strategies) investigated by Wu and Axelrod were Generous Tit-for-Tat (GTFT), which is like Tit-for-Tat, but when Tit-for-Tat would defect, 10 % of the time GTFT cooperates; Contrite Tit-for-Tat (CTFT), which cooperates after it accidentally defects; and Win-Stay, Lose-Shift (WSLS, which cooperates except if on the previous move it cooperated and the counterpart defected or it defected and the counterpart cooperated; this strategy is also called Pavlov in the literature). Creating four strategies from these three families, they added them to the 63 strategies in Axelrod's second IPD tournament and studied the effects of introducing small amounts of trembling (about 1%). GTFT won at 1% and CTFT did well, although their relative position was sensitive to the rate of trembling. In a subsequent ecological (or replicator dynamic) study, CTFT was the clear winner.

An early study (Marinoff 1992) re-considered the iterated Tit-for-Tat results of the Axelrod tournaments by investigating 20 strategies deliberately chosen to represent five "families" of strategies. Many of the strategies incorporated stochasticity, for example by cooperating with a certain probability under specified conditions. Using Axelrod's deterministic stage game payoffs, Marinoff investigated the rank robustness of the 20 strategies in all combinations (all ${}_{20}C_n,\ 1 \leq n \leq 20$ distinct tournament populations). In addition, Marinoff conducted tournaments on the ecological (or replicator dynamics) setup, also investigated by Axelrod. Across three measures of "robustness," and using the results of the two kinds of tournaments, Marinoff found Tit-for-Tat to be a generally good performer, but a noticeably poorer performer than several of the other strategies among the 20.

A recent experimental study (Fudenberg et al. 2010) examines play of IPD when there is stochasticity in the implementation of strategic choices (accidents, trembling, although the paper uses the term noise). The subjects reported using 20 different strategies. The study found substantial amounts of cooperative play by the subjects.

Finally, a recent major tournament (Rendell et al. 2010) investigated social learning in the context of a non-stationary multi-armed bandit. At each round of play, agents could choose one of three kinds of moves: Innovate (individual, asocial trial-and-error learning by manipulation of the environment); Observe (social learning, including copying of behavior by another agent); and Exploit (figuratively pull on one of the bandit's arms and, here only, be able to receive a reward). In all 104 entries were received and evaluated. The study found, among other things, that winning strategies invested heavily in social learning, at the expense of asocial learning. This general result was found to be stable in extensive robustness (or sensitivity) studies.

2.2 Desiderata for Strategic Choice

Our multi-criteria framework for strategy selection in games is closely related to the stream of literature on multi-agent learning in repeated games, where multiple desiderata are commonly considered for algorithmic comparison (Littman 1994; Bowling and

Veloso 2002; Greenwald and Hall 2003; Bowling 2004; Powers and Shoham 2005; Zawadzki et al. 2008). The history of this literature is largely that of proposing either a new algorithm that outperforms previous approaches on the same set of criteria, or a new collection of desiderata, together with an algorithm that performs well on these. As such, there is an implicit problem of strategy selection, that is, to guide one to choose a single learning algorithm from many options, which is an explicit aim of our work.

The criteria considered in this literature tend to be focused on the multi-agent learning context, and often are not generically meaningful. For example, the oldest criteria are convergence to a stage-game equilibrium in self-play and convergence to a best response against stationary opponents (Bowling and Veloso 2002), and newer desiderata that involve asymptotic no-regret learning guarantees (Bowling 2004). Any notion involving a stage-game solution concept, however, is fundamentally restricted to repeated games, and even convergence to equilibrium in self-play has been convincingly criticized in that context (Powers and Shoham 2005). Zawadzki et al. (2008) empirically compare a wide variety of multi-agent learning strategies on several classes of games using many of the desiderata proposed in the literature. Indeed, some of their criteria are closely related to ours. For example, they consider average reward in tournaments, just as we do. Most of the criteria they use, however, are specific to learning in repeated games (for example, no-regret criteria and convergence in self-play), and are, consequently, not generic.

3 Interdependent Security Games: Preliminaries

Imagine you live in a large condominium and there is a one-time opportunity to invest in fire-protection infrastructure. Specifically, for a price, you can have a sprinkler system installed in your apartment. If the sprinkler system is installed and a fire ever starts in your apartment (from whatever cause), you may assume that the system will reliably extinguish the fire and that any damage will be minimal. Your neighbors are also facing the same one-time decision. The complication in all of this is that even if you install a sprinkler system, if your neighbor does not and happens to experience a fire, the collateral damage to your flat will be the same as if you had not installed the sprinkler system and had had the fire yourself. Under what conditions would you elect to install the sprinkler system?

This is an example of what is called an interdependent security (IDS) game (Kunreuther and Heal 2003; Heal and Kunreuther 2005b). What is characteristic of these strategic situations is the possibility that a player may be harmed ("contaminated") through inaction by other players in investing in security, even if the player itself invests in security. When other players do not invest in risk reduction, the possibility of contamination (from realized risk events they experience) reduces the incentive for a player to invest for itself. Heal and Kunreuther (2005a) discuss the wide scope of applicability of these games, played either as one-shot games or as stage games in an iterated supergame. Examples include: airline security, industrial accidents (e.g., chemical plants, nuclear power plants), protecting buildings against terrorist attacks, protecting against risky behavior by other members and divisions of an organization (e.g., rogue traders, risk-seeking business units), and protecting networked computers from security violators.

Characteristic of IDS games is that the payoffs are given as lotteries (except when all players invest in protection). If you do not invest in a sprinkler you are exposed to the risk of a fire. If your neighbor does not invest, you are exposed to the risk of contagion damage if the neighbor has a fire. It is usual to represent games as having fixed and certain payoffs, rather than lottery payoffs. Standard solution techniques require this. Indeed, a main benefit of representing payoffs as utilities is that the decision maker is, in theory, indifferent between the expected value of a lottery whose outcomes are given in utilities and the lottery itself. If so, then any lottery outcome can be reduced to a certain outcome, its expected value. A similar situation often obtains when game models are applied in biology and the payoffs are given in fitnesses, usually defined as expected number of offspring. Roughgarden (2009) presents several nice examples. The main point for our purposes is that if payoffs are denominated in utilities or in expected numbers of offspring, then it is plausible to collapse lotteries into expected values.

And if not, perhaps not. Behavioral experiments (e.g., Kunreuther et al. 2009; Xiao and Kunreuther 2010) have indicated systematic differences in subjects' behavior between IDS games (with lottery payoffs) and their deterministic analogs (with lotteries replaced by their expected values). Setting the payoffs so that the games are Prisoner's Dilemmas (stochastic or deterministic), Kunreuther et al. find generally less cooperation in the stochastic IDS case than in the deterministic version of the game (Kunreuther et al. 2009). Even if we stick with conventional game theory, the fact that payoffs are realized through lotteries presents new strategic considerations, at least if we are modeling human players, whose choices may be affected. A player, for example, may choose to act in one way after experiencing the consequences of a security event and act in another way when the security event does not occur.

In sum, IDS games have lottery payoffs, model classes of phenomena of considerable import, are sources of anomalous behavior in subjects (at least from the point of view of classical game theory), and present challenges to policymaking (how to promote socially-useful investment?). In this paper we report on a computational strategic analysis of IDS games. Our goal is, when analytic methods are insufficient, to develop computational means and principles for understanding, and supporting intervention in, systems of strategically interacting agents. To this end, what follows reports on our exploratory development of such principles in the context of IDS games.

3.1 IDS Stage Game

The one-shot two-player IDS game, which is the stage game of the iterated interactions we study, represents the impact of externalities from player decisions about investing in security. Investing in security is costly, a fact that the model captures via a cost parameter c. Not investing in security results in a *direct* exposure to risk (we refer to the resulting loss as *direct loss*), that is, exposure due entirely to the player's own decision. We let p denote the probability of suffering a direct loss (say, due to a fire when a player chooses not to install sprinklers), with L denoting the actual (expected) magnitude of loss. Additionally, no matter what a player does, his counterpart's decision has an impact (an externality), but only if the counterpart does not invest in security. A

	Invest	Don't Invest
Invest	$-c$,$-c$	$-(c+pqL)$,$-pL$
Don't Invest	$-pL$,$-(c+pqL)$	$-(1+(1-p)q)pL$, $-(1+(1-p)q)pL$

Fig. 1 A stage game of the iterated IDS game

consequence of a player's decision not to invest is that his *counterpart* may suffer a loss *indirectly*, i.e., due solely to someone else's decision (we refer to the loss that results as *indirect loss* and denote its probability by q). However, a player only suffers an indirect loss if (a) he does not suffer a direct loss and (b) his counterpart *does* suffer a direct loss (again, thinking about a decision to install fire sprinklers, the fire spreads to a neighbor only if there is a fire at all, and only if the neighbor's apartment has not already burned down). The formal payoff matrix of an IDS stage game, with payoffs given in expected costs, is shown in Fig. 1. To connect the formalization with our description above, consider the payoff entry when both players player "Don't Invest", in which case each player's expected loss is $-pL-(1-p)qpL$. This loss reflects the underlying principle that "you only die once", that is, either a direct loss is incurred with probability p, or, when it is not incurred [with probability $(1-p)$], the indirect loss to the player is incurred only when his counterpart first incurs a direct loss (with probability p), and this loss is transferred to this player (with probability q).

The following is a complete characterization of the pure strategy equilibria of the IDS stage game:[2]

1. *Invest* is a (strictly) dominant strategy equilibrium iff $c < pL - p^2qL$;
2. *Invest* is a Nash equilibrium iff $c \leq pL$;
3. *Don't Invest* is a Nash equilibrium iff $c \geq pL - p^2qL$;
4. *Don't Invest* is a (strictly) dominant strategy equilibrium iff $c > pL$;
5. *(Invest, Don't Invest)* is a Nash equilibrium iff $c = pL$ and $q = 1$.

From item 5 above, the lone asymmetric pure strategy profile in the stage game is an equilibrium in only a very special case, so it suffices to focus on the symmetric equilibria.

Since the particular settings of all the parameters have considerable freedom from a purely game theoretic perspective, we chose to explore those values of c scaled roughly at the values commonly used in the IDS behavioral experiments. Specifically, we fix a "baseline" setting of $c = 45$, and a "baseline" setting of $p = 0.4$.

Additionally, we vary both c and p systematically in the intervals $[35, 60]$ and $[0.2, 0.8]$ respectively, each while keeping the other fixed at its baseline value. Throughout, L is fixed at 100 and $q = 0.5$. We note that given the baseline settings of c and p, the expected value IDS stage game becomes an instance of a Prisoner's Dilemma (Kunreuther and Heal 2003; Heal and Kunreuther 2005b).

Using the characterization above, we can arrive at the following *specific* single-stage equilibria in the parameter ranges that we consider:

1. When $c \in [0, 32)$, *Invest* is a dominant strategy equilibrium (fixing p at baseline);

[2] We do not deal with the mixed strategy equilibria of the stage game here, since pure strategy equilibria always exist in our setting.

2. When $c \in [32, 40]$, both *Invest* and *Don't Invest* are Nash equilibria (fixing p at baseline);
3. When $c \in (40, \infty)$, *Don't Invest* is a dominant strategy equilibrium (fixing p at baseline);
4. When $p \in [0, 0.45)$, *Don't Invest* is a dominant strategy equilibrium (fixing c at baseline);
5. When $p \in [045, 0.684]$, both *Invest* and *Don't Invest* are Nash equilibria (fixing c at baseline);
6. When $p \in [0.684, 1]$, *Invest* is a dominant strategy equilibrium (fixing c at baseline).

Our final point regarding the IDS stage game concerns efficiency of the decision to Invest. In general, the strategy profile (Invest, Invest) Pareto dominates (Don't Invest, Don't Invest) if and only if $c < pL + pqL - p^2qL$. Fixing p to its baseline value of 0.4 and setting the remaining parameters as above, this translates to $c < 52$; similarly, setting c to its baseline value of 45, gives the $p > \frac{3}{2} - \sqrt{1.35} \approx 0.338$.

3.2 The Iterated IDS Game

In our simulations, the stage game in Fig. 1 (for fixed settings of all parameters) is played a random number of times. Specifically, it is played at least once, and after every stage, it is iterated with a fixed probability $r = 0.968$, giving the expected number of rounds of ~ 31.

A major theoretical problem with iterated games is that the set of equilibria of these is extremely large: essentially, the entire range of "safe" (in the maxmin sense) to Pareto efficient strategic outcomes can be supported in equilibrium (Fudenberg and Tirole 1991). One classic approach is equilibrium refinement, attempting to rule out outcomes that fail some reasonable test (e.g., restriction to subgame perfect equilibria Fudenberg and Tirole 1991). However, it is not clear whether subgame perfection or any other refinement is sufficiently descriptive of human (or even computational agent) behavior. Instead, we restrict attention to a finite set of iterated game strategies (policies) which are loosely based on the strategies articulated by subjects playing iterated IDS scenarios in behavioral experiments.

Aside from complicating the structure of player strategies, iterated IDS games offer an opportunity to vary the *information* available to players about each other's past decisions. In particular, we make a distinction between *full* and *partial* information settings. Under full information, players observe their payoffs as well as all past decisions by their counterparts. Under partial information, players only observe their own payoffs, but have no direct information about counterpart decisions (although they can sometimes infer these from their payoffs). Additionally, our presentation of the stage game and the analysis of its equilibria above makes an implicit assumption that players only care about expected payoffs. In the iterated IDS game, we reveal actual realizations from the distribution of payoffs, rather than expected values. Thus, even if both players choose Don't Invest, there are likely rounds under which they do not incur any loss whatsoever, and in other rounds, they incur the full magnitude of the loss. Furthermore, players can condition their strategies on actual payoff realizations, rather than expected values. Our strategy space includes strategies that do just that. Note that in our setup there is a distinction between strategy sets in full and partial information settings: some strategies that are distinct under one strategic setup can be equivalent under another, and certain strategies can only be implemented in the full information setting (for example, Tit-for-Tat requires knowing the counterpart's decision in the previous round). We provide the full list of strategies and their descriptions in Appendix.

4 A Framework for Computational Strategic Analysis

Our central contribution is to introduce and illustrate a new *framework for computational strategic analysis* of iterated games which is motivated by Axelrod's tournaments, Axelrod and Hamilton's desiderata of evolution of cooperation, classical game theoretic analysis, and evolutionary game theory. (To our knowledge, this study is the first to combine elements of all of these elements). Our framework is defined specifically for two-player symmetric games [i.e., games in which two players share a common set of strategies S and common utility functions $u(s_i, s_{-i})$]. This focus is not in itself restrictive (there are natural ways to generalize the framework) but is primarily to simplify notation and discussion.

At the high level, we identify the following six desiderata that expand on and modify those proposed by Axelrod and Hamilton [these are also closely related to the desiderata proposed by Kimbrough (2012)]:

1. Tournament Robustness: Does the strategy thrive in a heterogeneous environment? We evaluate this with respect to a prior probability distribution over strategies in the player population.
2. Iterated Tournament Robustness: Does the strategy thrive when played against other high-quality strategies.
3. Stability: When played by all players, is the strategy stable in the face of unilateral deviations?
4. Strategic Resilience: Is the strategy resilient to variations in the population of strategies?
5. Initial Viability: Can a strategy gain a foothold in a non-cooperative environment starting with a small number of adopters?
6. Sensitivity Analysis: What are the values of predicted outcomes as a function of game parameters?

Our framework attempts to perform strategic analysis of two kinds: first, analysis of the relative merits of different strategies that may be adopted in the game, and second the impact of policy (e.g., public policy, as represented by specific game parameters) on the strategic predicament of the players and the resulting predicted strategic outcomes.

Our framework attempts to address many common criticisms of game theoretic analyses: it offers both a stability analysis of strategies, and an analysis of strategic performance of each specific strategy chosen by an individual in the face of a mixture of other strategies played by other individuals. Also, it yields a high-level analysis of the game made relevant to policy decisions. Finally, our use of multiple desiderata, each complementing the other, is targeted at multi-criteria decision making processes involved in the diverse situations in which strategy selection can be relevant, from an individual's choice of play in a game of interest (such as choosing an algorithmic trading strategy), to a policy maker attempting to incentivize a particular strategy which may or may not already be well represented in a population (such as cooperation in social dilemmas).

The key to creating a formal framework for strategic evaluation is formalizing all of its pieces, which we now proceed to do. First, let us abstractly define a stage game $\Gamma(z) = [I, S, u(s_1, s_2, z)]$ in which $I = \{1, 2\}$ is the set of players, S and u are the symmetric strategy set and utility function respectively. The utility function is defined both in terms of the joint strategies of players and a game parameter vector z (which we call a *game context*). Thus, our strategic evaluation will build in the impact of a particular game context (environment). We now proceed to formalize each part of the proposed framework.

4.1 Tournament Robustness

Axelrod's tournament has become an extremely important paradigm in assessing the relative value of iterated game strategies. It is thus a natural part of our framework. In the original tournament analyses, strategies were evaluated based simply on expected payoffs. The problem with such an evaluation is that we really wish to make a *relative* comparison: it is most informative to understand performance of each strategy as it compares to the best, rather than in absolute terms. For example, considering the expected payoff of a particular strategy is not in itself meaningful; what matters is how this payoff compares to the expected utility of the other strategies we are considering. This gives rise, below, to a *regret-based* definition, where regret, in general terms, is the amount by which a player would have preferred to play the best strategy, rather than the one in question. For example, if we measure the regret of a particular strategy s as being *low*, it means that s is extremely good: it is nearly as good as an optimal option for the player. In addition to being a relative metric (allowing us to evaluate each strategy in isolation), regret-based measures have the same interpretation regardless of the specific criterion: smaller regret is always better, since the strategy is closer to optimal. Consequently, we use regret-based measures for several other of our criteria as well.

Formally, define $V_T(s, z)$ as a *tournament value* of a strategy $s \in S$ in a game context z. This value is a measure of quality of a given strategy against some (prior) distribution over S (that is, over what opponents play). For example, V_T in Axelrod's tournament was the expected payoff against a uniform mixture of all participating strategies. We propose a slight variation and generalization of this approach. Specifically, let $F(\cdot)$ be a general distribution over S. We can view this as a prior distribution over all strategies in S; throughout, we use the uniform distribution as a running example. We then equate V_T with (the negative of) *payoff regret*, that is

$$V_T(s, z) = E_{s' \sim F}[u(s, s')] - \max_{t \in S} E_{s' \sim F}[u(t, s')] = \min_{t \in S} E_{s' \sim F}[u(s, s') - u(t, s')]. \tag{1}$$

In words, $V_T(s, z)$ is the difference between the expected payoff of s and the expected payoff of the best strategy t.

When we are dealing with finite strategy sets, the tournament value (payoff regret) is particularly easy to compute for all strategies. Specifically, fix z, let P be the full symmetric payoff matrix (empirical in our case), and let x be a vector representing probabilities over pure strategies in S. Thus, for example, when x represents a uniform distribution (as it does in our experiments below), $x_i = 1/|S|$, where $|S|$ is the number of strategies in S. The expected value vector EV of all strategies can then be computed as a matrix-vector product, $EV = Px$. Finally, we compute the vector

$$V_T = -[\mathbf{1}\max\{EV\} - EV],$$

where $\max\{EV\}$ is a the maximum value in EV and $\mathbf{1}$ is the unit vector.

4.2 Iterated Tournament Robustness

The original tournament, as conceived by Axelrod, measures robustness against an entire strategic pool, which may contain strategies that are truly poor and therefore unlikely to be played. We therefore consider a natural extension [anticipated by Axelrod and Hamilton (1981)] where we perform a tournament in two stages. The first stage proceeds to evaluate the payoff of each strategy against a prior distribution F with respect to the entire pool (say, uniform distribution, which is what we implement below). We then identify a fraction g of best performing strategies against F. More precisely, we rank the performance of all strategies in S, then choose $g|S|$ best strategies in terms of expected utility $E_{s' \sim F}[u(s, s')]$. Let S' be the resulting set of strategies. We next define $F' = F_{|S'}$, that is, a distribution over strategies F restricted to only choose strategies in S' with positive probability (the probabilities of $s' \in S'$ are thus renormalized). We then identify *iterated tournament value* $V_{T_{itr}}$ with negative tournament regret, just as in Eq. 1, but replacing F with F'. The intuition behind the regret-based definition of this measure is similar to that for tournament analysis above: we simply wish to know how much worse a candidate strategy is than the best possible (in a given consideration set), rather than what its expected payoff is in absolute terms.

4.3 Stability

Our next measure is that of strategic stability, which we refer to as $V_S(s, z)$, or *stability value*. In the context of symmetric games, it is natural to focus on the stability of *symmetric* strategy profiles, that is, strategy profiles in which all players play the same strategy. Thus, in symmetric games the strategic stability measure becomes a measure of quality of specific strategies, rather than full strategy profiles. Specifically, we let $V_S(s, z)$ be (the negative of) *game theoretic regret*, defined (for a symmetric profile s) as

$$V_S(s, z) = \min_{t \in S}[u(s, s) - u(t, s)].$$

The intuition behind the game theoretic regret measure is that we wish to assess how much a player gains from making an optimal decision, rather than choosing the strategy under consideration, *when the counterpart's decision is fixed*. When the strategy profile is symmetric, a strategy with low game theoretic regret means that when both counterparts choose it, neither has very much to gain to make a different choice. Another way to think about stability value is that it offers a generalization to equilibrium analysis: a symmetric Nash equilibrium has $V_S(s, z) = 0$, that is, there is, by definition, no alternative strategy s' which yields better payoff, as long as the counterpart's strategy is fixed at s. Since V_S is non-positive by definition, all Nash equilibria will have maximal stability value.

If the game is finite with payoff matrix P, it is particularly easy to compute a vector V_S, that is, (negative) stability regret vector of all strategies (more accurately, all symmetric pure strategy profiles):

$$V_S = -[\max\{P\}^T - \mathrm{diag}(P)],$$

where $\max\{P\}$ is a columnwise maximum of the matrix P.

4.4 Strategic Resilience

A strategy may be quite good in response to a fixed pool of others, but there is often much uncertainty about precisely which strategies are chosen by opponents. As such, a good strategy should perform well in tournament, but also be resilient to changes in the pool of strategies. We measure strategic resilience formally as follows. Let $\mathcal{S} = 2^S$ be the set of all subsets of S and let G be a distribution over $\mathcal{S}$. As above, let $F_{|T}$ be a prior distribution over strategies in S, restricted to only play those in $T \subset S$ with positive probability. We define *strategic resilience value* $V_{SR}(s, z)$ of a strategy s in context z to be

$$V_{SR}(s, z) = \min_{t \in S} \mathrm{Var}_{T \sim G}(E_{s' \sim F_{|T}}[u(t, s')]) - \mathrm{Var}_{T \sim G}(E_{s' \sim F_{|T}}[u(s, s')]).$$

In words, $V_{SR}(s, z)$ penalizes strategies s that have a high variance of expected payoffs with respect to random subsets of S. The actual measure is calibrated with respect to a strategy with minimal such variance (i.e., one that is most resilient to strategic variation by the opponent).

In our implementation we let G be a uniform distribution over all subsets in $\mathcal{S}$.

4.5 Initial Viability

In their work on evolution of cooperation, Axelrod and Hamilton (1981) argued that Tit-For-Tat was efficacious in promoting cooperation in part because it was able to survive even when there are relatively few individuals initially utilizing this strategy. More broadly, a policy maker interested in promoting a cooperative strategy that is not already widely deployed needs to ascertain whether such a strategy can ever viably take hold in a population and, if so, under what conditions. We can therefore view initial viability as imposing a constraint for a policy maker about a selection of strategies that can viably be promoted.

To quantify the idea of strategic viability (previously, this notion was largely qualitative), we must first quantify what it means for a strategy to "survive". To that end, we appeal to evolutionary game theory and, specifically, to *replicator dynamics* (Friedman 1991; Fudenberg and Tirole 1991). Since replicator dynamics are at the core of our approach to measuring initial viability, as well as performing sensitivity analysis below, we now describe (the discrete version of) it in detail.

Replicator dynamics begins with an initial population of agents, with each agent choosing a single strategy from the entire pool S. For each strategy s in this pool we can consequently quantify the fraction of agents, p_s, choosing that strategy:

$$p_s = \frac{\text{number of agents playing } s}{\text{total number of agents}}.$$

Starting with this initial strategy distribution p_s, replicator dynamics proceeds through a sequence of rounds. In each round, agents are randomly paired to play an instance of the game (in our setting, we mean the iterated game, rather than the stage game). We can compute the expected payoff for each strategy s in the population with respect to such random pairings; let this expectation for s be $Eu(s)$. Prior to starting the next round, the distribution of each strategy in the population, p_s is updated to be

$$p'_s \propto Eu(s),$$

and the process repeats. Replicator dynamics proceeds over L iterations, and frequently (in practical experience) it converges to a symmetric mixed strategy Nash equilibrium of the game, so long as every strategy in S has some representation in the initial agent population (i.e., $p_s > 0$ for all $s \in S$).

Now, suppose that we seed replicator dynamics with some probability distribution over strategies in S in which the target strategy s is played with probability p_s; let probabilities of other strategies in the seed distribution be denoted by p_{-s}. Next, run replicator dynamics until it either converges or reaches a limit L on the number of rounds (we use 200 rounds in our simulations below); let $N(z, p_s, p_{-s})$ be the final mixed strategy (population proportions) reached by replicator dynamics. We say that survivability of strategy s is high if its probability under $N(z, p_s, p_{-s})$ is significantly above 0.

To quantify how viable s is when it is underrepresented in the initial strategy pool, we let $p_s = f \min_{t \in S \setminus s} p_t$, where $f \in [0, 1]$; that is, s is initially played with probability that is some fraction of the smallest probability with which any other strategy is played. In our experiments below, we used $f = 1/2$ and used a uniform distribution over all strategies other than s in the seed distribution of replicator dynamics. So, for example, if there are 10 other strategies (11 strategies in total), every strategy is played with probability ~ 0.095, while the initial probability of s is ~ 0.048.

4.6 Sensitivity Analysis

One of the main reasons social dilemma games in general, and IDS games in particular, are of broad interest is that their analysis may enlighten policy. There has been little attempt, however, either in tournament-style analyses, or other game theoretic treatments of social dilemma games, to understand the impact of interventions (via manipulation of game parameters) on predicted outcomes. The goal of this section is to provide a general framework for doing such assessment by means of careful sensitivity analysis in the space of policy relevant parameters of the game model.

Recall that we denote by z a collection of parameters that influence the payoff functions of players in the game. In IDS games, parameter z could incorporate a subsidy that helps defray some of security investment costs; equivalently, it could incorporate the investment cost parameter itself, since a subsidy r to an initial cost c simply shifts the investment cost to $c - r$. The goal of sensitivity analysis is to characterize the outcomes of the game (e.g., in terms of Nash equilibria) as a function of z.

Formally, let $N(z)$ be a Nash equilibrium of $\Gamma(z)$ and define a measure $\mu(N(z), z) = \mu(N(z), u(N(z), z))$ which quantifies an *external value* (e.g., a policy value) of an equilibrium. For example, in IDS games such external value would be social utility, and specifically in Prisoner's Dilemma situations $\mu(N(z), z)$ would correspond to the fraction of player decisions that are cooperative. Alternatively, $\mu(N(z), u(N(z), z))$ may simply be the social welfare in equilibrium, which in a symmetric equilibrium of a symmetric game can be analogously quantified by $u(N(z), z)$. When we consider μ with respect to some distribution over $N(z)$, we use the simpler notation $\mu(z)$, and keep the distribution itself implicit (a specific manner to arrive at a distribution over $N(z)$, which need not be analytic, is explicated below).

To perform the quantitative sensitivity analysis of policy alternatives in terms of game parameter vector z, we map out $\mu(z)$ as a function of z for one or more such *external value* measures μ. Thus, for each $\Gamma(z)$:

1. For $k = 1, \ldots, K$:
   - Sample a mixed strategy m_k uniformly randomly from the unit simplex (i.e., the set of all mixed strategies) over S;
   - Seed replicator dynamics (Friedman 1991; Fudenberg and Tirole 1991) with m_k as the initial distribution of strategies S in the population;
   - Run replicator dynamics until it either converges or reaches a limit L on the number of rounds (we use 200 rounds in our simulations below); let $N(z, m_k)$ be the final mixed strategy (population proportions) reached by replicator dynamics starting at m_k;
2. Define $\mu(z) = \frac{1}{K}\sum_{k=1}^{K} \mu(N(z, m_k), z)$, that is, the sample mean of the policy value based on K sample runs of replicator dynamics.

Note that the final step of the framework resolves the equilibrium selection issue computationally by producing a distribution over equilibrium outcomes by iteratively sampling final outcomes of a replicator dynamics evolution.[3]

[3] We are indebted for this idea to Walsh et al. (2002). It assumes that replicator dynamics converges, which it did in every instance we had observed.

In what follows we concretely illustrate our framework by using it to offer some insights about the strategic landscape of iterated IDS games, as well as a policy-relevant measure motivated by actual IDS concerns.

5 Computational Strategic Analysis of Strategies in iterated IDS Games

In this section we perform computational strategic analysis, following our framework, of three classes of strategies in iterated IDS games. The first class involves strategies that either punish or reciprocate the behavior of the counterpart, and includes the much touted Tit-for-Tat. The second class includes the two strategies which prescribe that a player either always or never invest in security. The third class of strategies focuses on the player's ability and desire to act based on observed realizations of negative events, either choosing to invest, or not, in response to them. The first class of strategies only applies in the complete monitoring setting, whereas the latter two classes are studied both when there is full and partial feedback about the counterpart's decision in the previous round.

5.1 Computational Setup

We ran 500 simulations of iterated IDS instances using a NetLogo implementation[4] for a collection of game parameter settings and used the data to construct empirical payoff matrices of the iterated game (that is, average simulated payoffs of joint iterated game strategies). Additionally, we obtained for each joint strategy of both players the corresponding value of μ (as described below). We explored a fixed grid of c and p values, with $c \in \{35, 40, 45, 50, 55, 60\}$ and $p \in \{0.2, 0.4, 0.6, 0.8\}$. When we varied c, we maintained p at its baseline value of 0.4 and, similarly, we varied p while maintaining $c = 45$. We present our results below in two-dimensional $\{c, p\}$ space mapped out by the above parameter values. For each c and p, we plot a square box; lighter color of this box means that the strategy under consideration performs better for this setting of game parameters.

We partition the space of c and p values above into three significant regions:

- The first (upper left, corresponding to low investment cost and/or high probability of direct loss) is the *cooperation region*: in this region, investing in security is both Pareto dominant and a Nash equilibrium (but not necessarily a unique equilibrium).[5]
- The second (lower right, corresponding to high investment cost and/or low probability of direct loss) we term the *defection region*, since not investing in security

[4] The NetLogo implementation used for our data can be found at http://opim.wharton.upenn.edu/~sok/netlogo/IDS-experiments.nlogo. An updated version can be found at http://opim.wharton.upenn.edu/~sok/AGEbook/nlogo/IDS-2x2-Tournaments.nlogo.

[5] We remind the reader that although IDS games have stochastic payoffs, and the behavioral experiments have shown that this matters to players in laboratory studies, our discussion here proceeds in terms of the estimated expected values realized by our computational experiments.

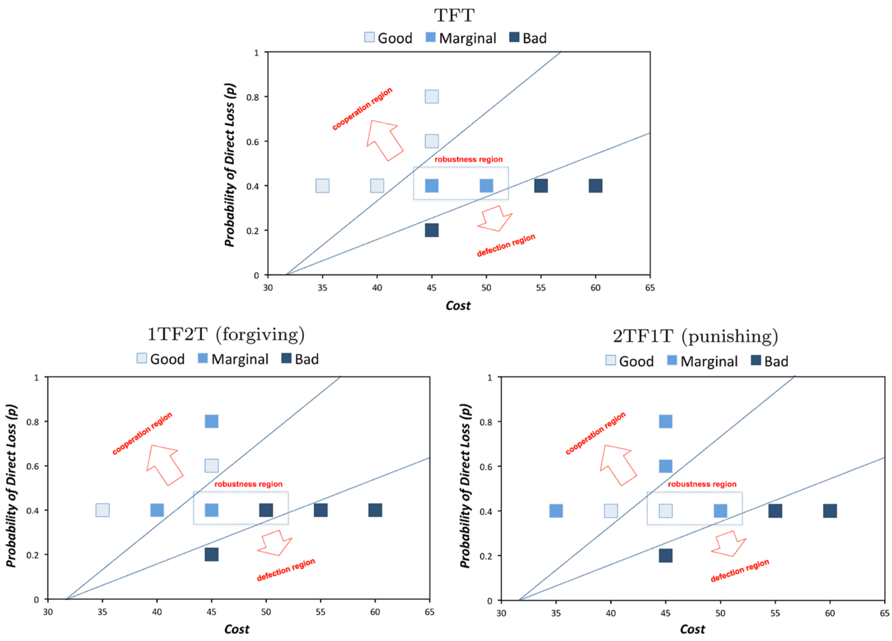

Fig. 2 Relative tournament performance (as measured by V_T) of the three punishment and reciprocity strategies. *Lighter boxes* correspond to lower "tournament regret" and, therefore, *better* performance in tournament relative to the best strategy

is both a dominant strategy Nash equilibrium and a Pareto optimum for this set of c and p values.

– The third, intermediate, region is the *robustness region*. In this region, security investment is Pareto dominant, but is also strictly dominated by non-investment in the IDS stage game; in other words, these settings of c and p make the IDS game a Prisoner's Dilemma.

5.2 Strategic Punishment and Reciprocity

One of the most significant findings in the original work on tournaments in Prisoner's Dilemma games by Axelrod was that Tit-for-Tat (TFT) was the top performer. An extremely appealing aspect of this strategy is its interpretation as exhibiting strategic punishment (when the counterpart defects) and reciprocity (as long as the counterpart cooperates). One can, in principle, define a class of strategies with similar structure, where one can be more forgiving of defections, or more punishing. We consider three examples from this class: Tit-for-Tat, 1-Tit-for-2-Tats (more forgiving; we abbreviate it as 1TF2T), and 2-Tits-for-1-Tat (more punishing; abbreviated as 2TF1T), which we compare below based on the five metrics identified in our general framework.

Tournament Performance Our first comparison is that of tournament performance of the three variants of strategic punishment and reciprocity strategies, shown in Fig. 2.

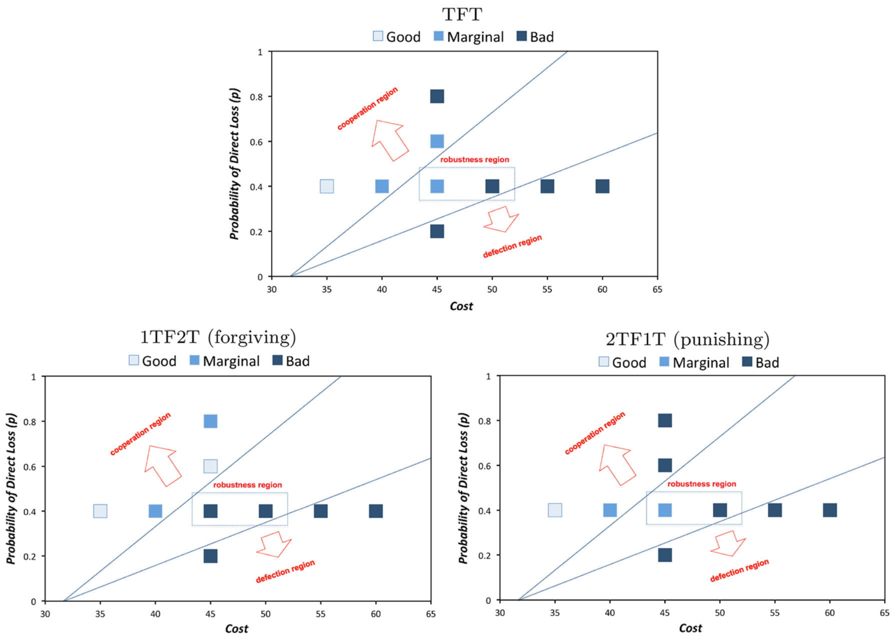

Fig. 3 Relative performance in a tournament of punishment and reciprocity strategies in which only the top 1/3 strategies participate (based on performance in the "standard" tournament). Measured by $V_{T_{itr}}$, *lighter boxes* correspond to lower "iterated tournament regret", or *better* performance as compared to the best

First, notice that there is no distinction between these strategies in the defection region: all perform quite poorly. This observation agrees with intuition: when defection is clearly superior, there is no sense in doing anything else (e.g., reciprocating cooperation). In the cooperation region TFT is superior to the others. The punishing strategy is clearly worse because investing is both a Pareto optimal and an equilibrium strategy here. The forgiving strategy is particularly vulnerable when there are two equilibria, since it doesn't adjust to play the defection equilibrium as quickly (e.g., when the counterpart Never Invests).

A surprise emerges when we inspect the robustness region: here (in the baseline setting) the punishing strategy is superior to TFT, which is better than the forgiving variant. Recall that this region corresponds to a Prisoner's Dilemma payoff structure. It thus appears that TFT (in its traditional incarnation) is not sufficiently punitive in certain Prisoner's Dilemma incarnations. For the strategic pool we consider, it is therefore at a greater disadvantage compared to defecting strategies (e.g., Never Invest) when paired against consistent cooperators who occasionally defect with some fixed probability.

Iterated Tournament Performance Our next step is to measure the strategies in terms of their iterated tournament performance.

The results for the class of strategies involving punishment and reciprocity are shown in Fig. 3. Interestingly, the picture that emerges here is somewhat different than the one observed under the classical tournament analysis above. The first difference is that here the forgiving strategy is actually the best in the cooperation region. The reason

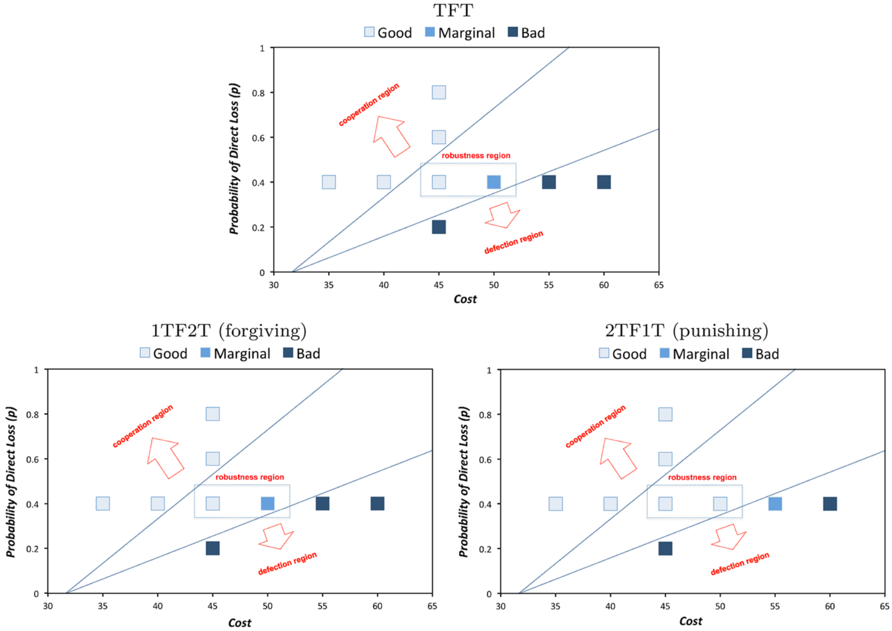

Fig. 4 Stability (V_S) of symmetric equilibria corresponding to the punishment and reciprocity strategies. *Lighter boxes* correspond to higher stability, meaning that a player has less to gain by changing his strategy when the counterpart's strategy remains unchanged

is that iterating the tournament even once eliminates many defectors in this region; now it pays to forgive. The second difference is that now TFT and the punishing variant are no different in the robustness region, though both are still better than the forgiving strategy. Indeed, both TFT and 2TF1T are now *worse* ("marginal"): it appears that initially iterating the tournament promotes greater defection in the robustness region.

Equilibrium Stability Equilibrium stability is a fundamentally different measure than those based on the tournament. Here, we are concerned purely with stability of a strategy to deviations; in other words, we favor a strategy which, when played by the two counterparts, is nearly the best one for both.

We present the results of stability analysis for the punishment and reciprocity strategies in Fig. 4. All strategies are actually rather similar in that regard, except in the robustness region, where the punishing variant is more stable than the others; 2TF1T appears quite stable (nearly a symmetric Nash equilibrium) in the entire robustness region.

Strategic Resilience Figure 5 compares strategic resilience of the punishment and reciprocity strategies. In the cooperation region, the forgiving strategy is the most resilient to opponent variation, slightly more so than TFT, while the punishing strategy is the least. This echos our observation for the iterated tournament measure. It's more resilient than both alternatives in the region where investing in security is a dominant strategy, since cooperation is a best response no matter what the opponent plays, so

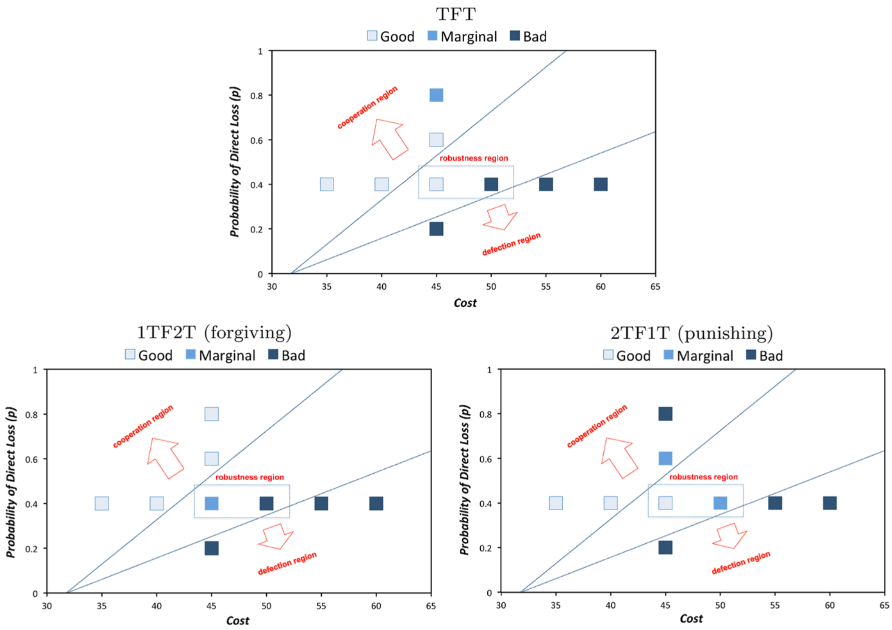

Fig. 5 Strategic resilience (resilience to variation in opposing strategies) of the punishment and reciprocity strategies, as quantified by $V_{SR}(s, z)$. *Lighter boxes* correspond to higher resilience

being forgiving means greater propensity to invest even when the counterpart does not. The forgiving strategy is superior to the punishing variant even when there are two equilibria likely because punishment may push the counterpart to play the socially inferior equilibrium. The three punishment and reciprocity strategies are identical in the defection region, but the punishing strategy is the most resilient in the robustness region, just as we had observed using most of the other measures of strategic quality.

Initial Viability Our final measure of strategic quality is whether a strategy can survive when initially underrepresented in the population. Figure 6 explores this initial viability property as it pertains to the punishment and reciprocity strategies. The main observation here agrees with the previous measures: the punishing variant of TFT is not very viable in the cooperation region (likely for similar reasons), but is better than the other variants in the robustness region.

Summary Our metrics of strategic quality are in broad agreement about relative efficacy of the three punishment and reciprocity strategies that we study. Here we summarize our main observations:

– The punishing strategy is generally the best in the robustness ("Prisoner's Dilemma") region, particularly as defection yields greater gain due to the high investment cost if one cooperates;
– The three strategies considered in this section are all quite poor in the defection region, while the classic TFT as well as its forgiving variant are generally good in the cooperation region;

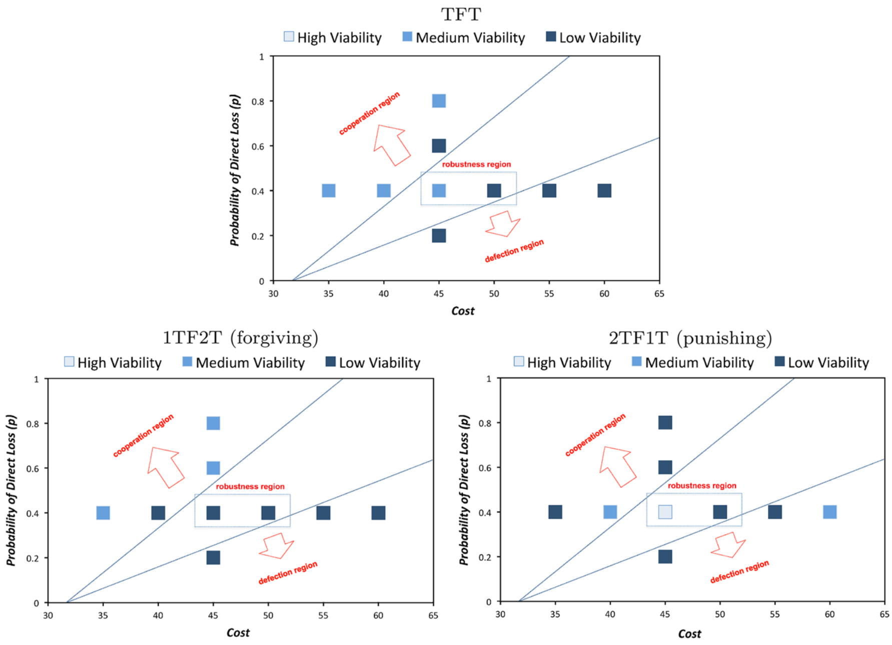

Fig. 6 Initial viability of punishment and reciprocity strategies

– A multi-stage tournament is unfavorable to the TFT variants: for some values of joint parameters, even in the cooperation region, none of these is close to optimal (with mostly cooperative strategies left after the first tournament round, ability to punish seems to be somewhat disadvantageous here). Indeed, even in the region of parameters corresponding to the widely studied "Prisoner's Dilemma" (we term it the robustness region), TFT can be among the *worst* performers on this criterion;
– TFT and related strategies are weak on the Initial Viability metric even in the cooperation region. This suggests that effectiveness is likely to be quite limited if the initial strategic pool does not already include a significant amount of cooperation, which is precisely the context where policy would likely be deployed (after all, if cooperation is already relatively dominant, there is no need to institute any intervention). The most effective policy intervention would therefore first provide incentives to significantly increase the tendency to invest in security (for example, making it a dominant strategy), and the incentives may subsequently be reduced once a sustainable level of cooperation is reached. This situation bears similarity to introduction of new and superior technologies (e.g., QUERTY vs. DVORAK keyboard layout) where an inferior technology already has significant network effects; the first step is to overcome network effects, at which point the better technology can sustain itself (Rogers 2003).

5.3 Consistent Strategies: Always and Never Investing

In this section we consider the two simplest IDS strategies: Always and Never Invest. Both of these play the corresponding one-shot strategy in every round of the game,

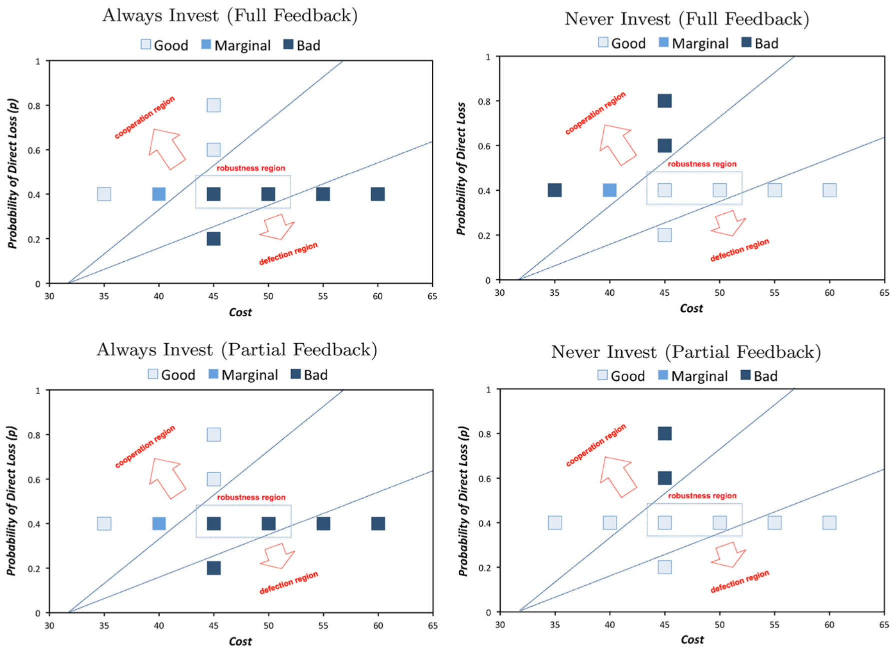

Fig. 7 Relative tournament performance (as measured by V_T) of Always and Never Invest strategies. *Lighter boxes* correspond to lower "tournament regret" and, therefore, *better* performance in tournament relative to the best strategy

hence our use of the term "consistent". They are also, in a sense, the two extremes: always invest is an extreme example of cooperation, while Never Invest is the ultimate defection strategy. These strategies are well defined in both full and partial feedback setting; we therefore compare their performance in both of these cases.

Tournament Performance We begin again with the comparison of Always and Never Invest strategies based on the tournament. The results, shown in Fig. 7 suggest that Never Invest is clearly superior to Always Invest in every setting where defection is a dominant strategy (the defection and robustness regions); indeed, it is nearly the best strategy in these regions. The likely reason is that Never Invest gains more from exploiting weak strategies (either non-adaptive, such as Always Invest, or those that adapt poorly, such as investing or not investing after a loss) than it loses (relative to, say, TFT) by not cooperating with adaptive and cooperative strategies. Matters are somewhat more interesting in the cooperation region. While Always Invest is highly effective through most of this region, the relative quality of Never Invest depends on whether the setting involves full or partial monitoring. Under full monitoring, Never Invest is quite poor in the cooperation region; when there is only partial monitoring, however, Never Invest is nearly optimal in a large fraction of it ($c \leq 40$, when $p = 0.4$). The key reason for this distinction is that under partial monitoring, cooperation-promoting strategies which are based on punishment and reciprocity (e.g., TFT) cannot be implemented when we don't know whether the counterpart invested or not. Thus, defection remains a highly efficacious candidate.

Iterated Tournament Performance Iterated tournament comparison is quite favorable to Never Invest: it is now highly efficacious in nearly all settings (in both full and partial feedback). Always Invest also benefits, albeit slightly: it now exhibits uniformly high quality in the cooperation region.

Equilibrium Stability In terms of equilibrium stability, partial feedback is unfavorable to Always Invest, and favorable to Never Invest. The reason is that, as we noted above, absence of punishment and reciprocity strategies removes high quality alternatives to defection, making it more efficacious across a broad array of IDS game structures.

Initial Viability The initial viability results broadly support the main observation of this section that partial feedback promotes defection, as compared to full feedback in the robustness and defection regions: under partial feedback, Never Invest emerges as a significant player in the strategy population even when significantly under-represented at the beginning across a wide array of IDS parameters.

Summary The central finding in this section is that partial feedback promotes defection, making the Never Invest strategy more viable and effective across a broad array of game parameters. The basic problem encountered when players cannot monitor each other's actions is that there is little upon which to base reciprocity and punishment. When selfish, but socially damaging, actions are not observable, it is no surprise that a much larger proportion of the population will perform such actions. Ostrom (2009) offers examples of effective "self-organized" systems for managing commons, where perpetrators of asocial actions are punished by the community. For example, fishing fleets in Gloucester, MA developed a system of rules and penalties, which become incrementally more severe for repeated offenses (for example, after a third offense, lobstermen break the trap of the fisherman who breaks the rules). Such rules and penalties are only communally enforceable if the rule breaking actions are *observable*, that is, correspond to our notion of "monitoring". In the case where it's difficult to observe asocial actions, Ostrom observes that self organization is far more difficult to achieve, and "defection" prevails. In such a situation, the policy maker's options are either to significantly increase the relative payout of socially desirable actions (e.g., through subsidies), or to increase the means of monitoring activity (for example, increase the size of the police force).

5.4 Loss-Conditional Strategies

Our final set of strategic analyses pertains to strategies that appear to be closely related to human behavior in stochastic settings. People tend to react strongly to a recent negative event, a behavior that is a special case of the availability heuristic (Camerer and Kunreuther 1989). In our case, the negative event is a loss, be it due to a direct or indirect exposure. The two possible reactions in IDS games are to invest (giving rise to the invest after loss strategy) and not to invest (yielding the don't invest after loss strategy). The don't invest after loss strategy has partial resemblance to Tit-for-Tat: a loss can be viewed as an indication of non-cooperative (non-investment) behavior by the counterpart, and the appropriate punishment is to stop investing.

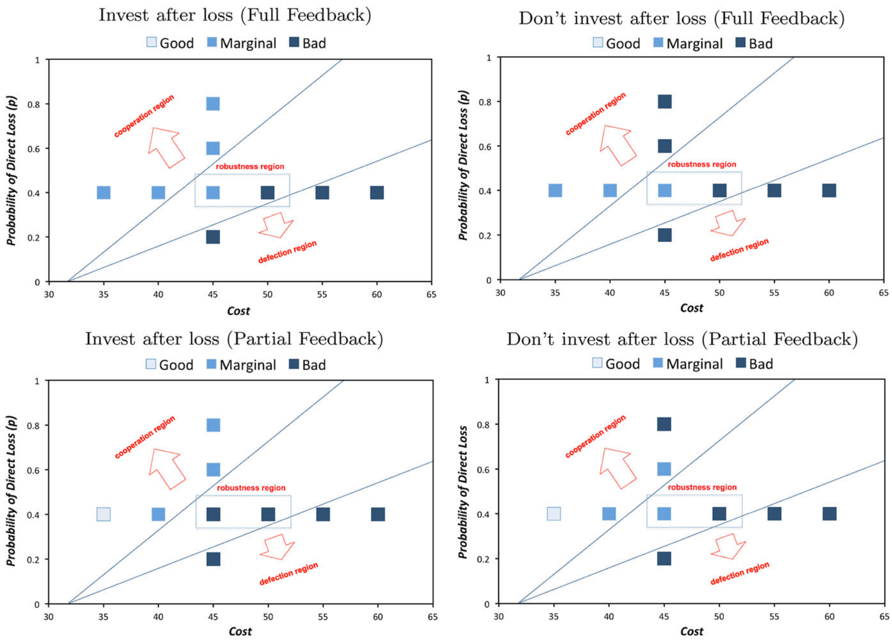

Fig. 8 Relative tournament performance (as measured by V_T) of the two strategies that invest and don't invest after loss, respectively. *Lighter boxes* correspond to lower "tournament regret" and, therefore, *better* performance in tournament relative to the best strategy

Tournament Performance Our findings regarding the two loss-conditional strategies (invest and don't invest after a loss) are highly consistent across the different measures (with the exception of viability, to which we turn presently), and we therefore focus only on their performance in the tournament.

Figure 8 paints a very negative qualitative picture of these two strategies. First, observe that there is little qualitative difference between the two strategies: both are very poor in the defection region, and both are marginal or poor in the rest of the $\{c, p\}$ landscape. Investing after a loss seems to be somewhat better than not in the cooperation region, a rather intuitive finding since here cooperation (investing) is indeed a superior strategy. Overall, however, neither strategy is very close to optimal throughout the entire landscape, save a couple of exceptions when investment costs are lowest and only partial feedback is available, in which case investing after a loss does well. The poor performance of these strategies is rather significant, and quite surprising. In the partial feedback setting, not investing after experiencing a loss is reminiscent of TFT, as we noted above, and yet its efficacy does not come near that of TFT. The reason is that the response to a loss is not sufficiently correlated with a counterpart's decision: a loss could be either direct or indirect, and, moreover, the counterpart may well not invest for a long time before he is discovered when a loss is finally experienced. Moreover, these being strategies that are similar to those employed by humans in analogous situations, our results have important policy implications: a policymaker interested in socially desirable outcomes should use information campaigns to steer

Table 1 Qualitative performance of the strategies and strategy classes that we have studied in depth. "++" indicates near-optimal performance on all criteria; "+/ + +" means near-optimal performance on all but one criterion in nearly the entire region; "+" means near-optimal performance on most criteria in most parts of the region; "0" means near-optimal on some criteria in some portion of the region, and not on other criteria; "−" means poor on most criteria; "− −" means among the worst on all or nearly all criteria

Strategy name	Cooperation		Robustness		Defection	
	Full feedback	Partial feedback	Full feedback	Partial feedback	Full feedback	Partial feedback
Always Invest	+/ + +	+/ + +	− −	− −	− −	− −
Never Invest	0	0	+/ + +	++	++	++
Punishment/reciprocity (TFT-variants)	0	N/A	0	N/A	− −	N/A
Loss-conditional	−	0	−	−	− −	− −

people away from such highly exploitable strategies. (Indeed, somewhat in contrast with many previous studies of IPD, all of the strategies we consider are actually based on observed behavior in actual IDS experiments.)

Initial Viability Initial viability offers an entirely one-sided story for the loss-response strategies: neither is viable when initially underrepresented in the population, in any of the settings we studied. Given the extremely poor quality of these strategies described above, this is actually good news: if a policy can shift behavior away from adopting such strategies initially, it is unlikely they will take root in the future.

5.5 Best-Performing Strategies

Overall performance of the different strategies and classes of strategies we have considered is shown schematically in Table 1. We now summarize the best performers from these classes in each of the three regions of the IDS parameter space.

Cooperation Region With the exception of initial viability, TFT and/or at least one of its derivatives, as well as Always Invest, tend to be near-optimal in the cooperation region on all metrics when there is full feedback (there are a few qualifications to that, which we described and explained above). With partial feedback, Never Invest becomes near-optimal as well in a portion of that region. The reason for that is two-fold: Never Invest is also an equilibrium in this region, and strategies involving punishment and reciprocity become difficult to support, removing a strong barrier to defecting strategies.

Robustness (Prisoner's Dilemma) Region In the robustness (Prisoner's Dilemma) region, Never Invest is in every case among the top performers (although not necessarily the optimal strategy). While variants of TFT are often in the near-optimal class as well, this is only the case for a subset of settings and criteria; indeed, their tournament performance is not among the (near-)optimal strategies, although it is often near-optimal in the iterated tournament setting (previous tournament comparisons which yield TFT as the best strategy generally include relatively sophisticated alternatives, whereas our tournament includes a number of relatively bad strategic choices, such as investing after a loss, which are also highly exploitable by Never Invest).

Defection Region In the defection region, Never Invest is both a stage game dominant strategy and a Pareto optimum. Thus, while Never Invest is always among near-optimal strategies in this setting, none of the other strategies that we studied in depth here are near-optimal.

6 Propensity to Invest in Iterated IDS Games

From a policy perspective, a key reason to study IDS games is to explore the impact of decision externalities among players and consider what policy tools can be effective at bringing about socially superior outcomes. Over a broad range of parameters of IDS games, invest decisions by both players Pareto dominate all the other pure strategy profiles. However, investing may be against individual self-interest, as it is when IDS parameters indicate that "Not Invest" is a dominant solution, as is the case in social dilemmas when outcomes are deterministic. A natural question for a policymaker is how to incentivize socially desirable outcomes through minimal intervention.

Two useful incentives in the policy toolbox are subsidies on socially desirable decisions, or fines for socially undesirable ones. Concretely, in IDS games let us suppose that we wish to incentivize players to invest in security, at least where investment outcomes are Pareto optimal. For simplicity, we suppose that the policymaker offers a subsidy r to those who choose to invest. The effect of this subsidy on the game structure is to lower the effective investment cost from c to $c - r$. Equivalently, rather than studying the impact of r on outcomes directly, we proceed to engage in sensitivity analysis of game outcomes, in terms of propensity of players to invest (cooperate), in terms of the cost of investing in security, c.

Formally, let f be the expected fraction of players to invest in the game equilibrium reached by replicator dynamics (following the process we outlined in Sect. 4). We wish to get a handle on the function $f(c)$, fixing p to its baseline value of 0.4. The goal is to choose c such that $f(c)$ is high enough. On the other hand, we wish to make r as small as possible, since subsidies are costly.

Another question of interest to a policymaker is the impact of the probability of direct loss, p, on the propensity of players to invest, $f(p)$. In terms of actual policy implications, the policymaker may be able to impact the *perception* of p by the agents engaging in security decisions, by making the possibility of a loss more salient to them (e.g., through advertising).

In addition to studying investment behavior as a function of parameters c and p, of vital interest to policy is the impact of monitoring (in our terminology, full and partial feedback) on the number of agents who invest in security, as well as on the efficacy of policy instruments such as fines and subsidies. Thus, we present $f(c)$ and $f(p)$ for both full and partial feedback settings below.

6.1 Impact of Subsidy on Investment

We now present one of our main results: the impact of increasing investment cost on propensity of players to invest in security. We quantify this propensity as the

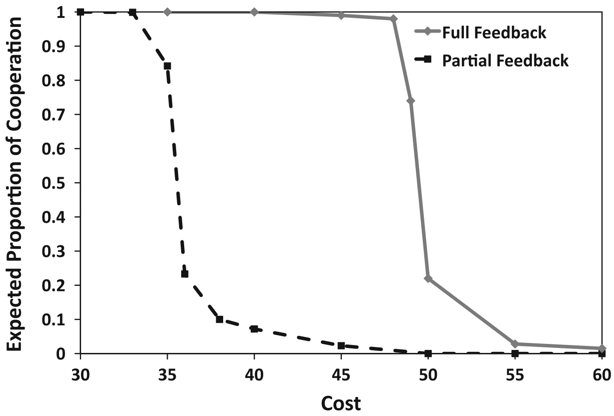

Fig. 9 Proportion of cooperation versus cost in full and partial information games

expected fraction of investors in the long run of replicator dynamics, averaged over runs seeded with a uniformly random initial distribution of strategies (described in detail in Sect. 4.6). Figure 9 shows several provocative results. First, the gap between full and partial feedback is dramatic. At the highest cost (c = 60) both partial and full feedback conditions lead to Not Invest. However, in the full feedback condition when $c < 50$ most players Invest even though Not Invest is a dominant strategy equilibrium until $c \geq 40$ and is a Nash equilibrium if $c > 32$. When there is partial feedback, only when $c = 35$ do most players follow the strategy Invest. From a policy perspective, it may well be more cost effective to regulate and enforce a high level of monitoring than incentivize investment in the partial feedback setting.

The second surprising observation is that investment levels do not increase gradually as investment cost falls. Instead, we have what appear to be phase transitions at $c = 50$ for full feedback and at $c \approx 35$ for partial feedback. This is quite significant, since it suggests that small subsidies or penalties may have either no impact at all, or cause a sea change.[6]

One reason that subsidies do not necessarily help in the IDS game we study is that the game involves only two players, both identical (homogeneous). When players have heterogeneous preferences, and these are common knowledge, subsidizing a few carefully selected players (e.g., those with smaller investment costs) can tip the others to invest. As an example, if one considers investing in a sprinkler system in a large building, this positive impact of this investment may be almost negligible if no one else is investing, since the fires originating elsewhere will still do significant damage to the unit. On the other hand, when nearly all others (particularly, close neighbors) have also installed sprinklers, their joint operation is likely to stop fires from doing much damage before spreading, and the marginal impact of installing a sprinkler may now be well worth the investment. With heterogeneous preferences, it may be far easier to incentivize tenants whose preferences are already easy to tip towards

[6] Note that the observed phase transitions are not immediate from stage game analysis, since the phase transition points *do not correspond* to the stage game transitions between equilibria.

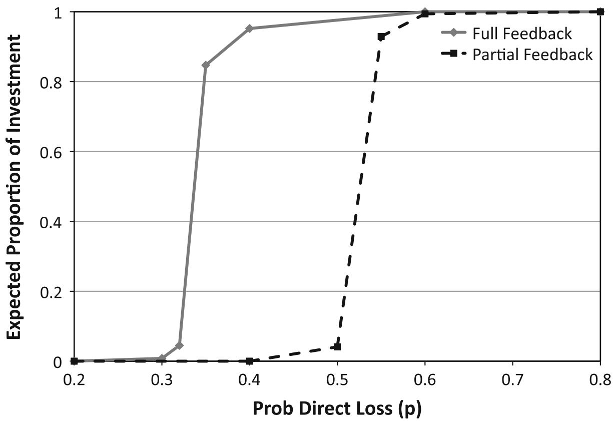

Fig. 10 Proportion of cooperation versus cost in full and partial information games

investing. Since the investment by these players may make investment worthwhile for others, subsidizing these tenants so that they invest may be quite effective in ultimately inducing a socially desirable outcome. As an another example of the importance of heterogeneity in preferences, it could be that several tenants are far more likely to start a fire than others, and by targeting specifically these tenants, a policy maker may achieve a higher investment level overall at a lower total expected costs. (See Heal and Kunreuther 2005b for further discussion).

6.2 Impact of Probability of Direct Loss on Investment

Our final consideration of this section is an analysis of propensity to invest as a function of direct loss probability p. Figure 10 shows the results, which take a familiar form: there is a dramatic difference between full and partial monitoring settings, and in both there appears a distinct phase transition from non-investment to full investment levels. For policy purposes, regulating the level of monitoring, whenever feasible, seems again of vital importance. Additionally, the sharp boundaries suggest that small differences between particular strategic scenarios in which the players may find themselves can make a dramatic difference. Thus, it may even pay to slightly alter peoples' *perceptions* of the probabilities of direct loss. As an example, it is known that adding detail into a description of an event makes it appear more likely (Camerer and Kunreuther 1989). Shifting focus to losses, and adding details to their descriptions, may therefore make it sufficiently vivid to push the population towards significant investment in security.

7 Discussion and Policy Implications

We presented an extensive framework for strategic and policy analysis of general games, with special focus on social dilemma scenarios. As an illustration, we chose an important setting involving security decisions with externalities, commonly referred to as IDS games. IDS games characterize a variety of strategic predicaments, including a

Prisoner's Dilemma. Therefore, considering various combinations of parameter values provides us with general insight about this class of games both from a strategic and public policy perspective.

Our first set of results pertains to the relative efficacy of strategies across the strategic IDS settings we considered. Broadly, we found that Tit-for-Tat, while reasonably effective in a variety of settings, is not quite as robust in social dilemma situations as its more punishing variant, 2-Tits-for-1-Tat. We observed, additionally, that strategies that condition decisions on a recently observed loss are almost universally poor choices.

One of our most significant results from a public policy perspective is the importance of monitoring past decisions of a counterpart. When monitoring is allowed (as it was in the original Axelrod tournaments, and in most follow-up work), cooperative strategies perform significantly better than in the partial feedback (no monitoring) settings. A strategy like Tit-for-Tat cannot be directly implemented when one cannot monitor opponent actions, and strategies that respond to a loss perform quite poorly. This implies that when agents are uncertain as to what action their counterparts have taken one may need to consider well-enforced regulations and standards to obtain cooperation. Alternatively, subsidizing or fining key players to induce them to cooperate may lead others to follow suit as shown by Heal and Kunreuther (2005b).

Our sensitivity analysis, which mapped out an expected fraction of security investment in the population as a function of investment cost c and the probability of direct loss p corroborates the significance of monitoring: a remarkably large gap exists between the fraction of investors under full and partial monitoring settings. Indeed, the response to possible policy interventions such as subsidies on investment decisions (intervention that lowers c) or attempts to make the possibility of loss due to non-investment (intervention that increases perceived p) exhibits a phase transition: in the vast expanse of the parameter regions, even substantial changes to either c or p have almost no impact. However, if the current social dilemma features happen to fall near a transition boundary, even a small intervention can have dramatic impact. In contrast, increasing the level of monitoring, perhaps through direct regulation, is likely to effect a dramatic impact on the level of security investment over a broad space of strategic IDS predicaments.

While our application was restricted to two-player homogeneous (symmetric) IDS games, our framework is more general. Indeed, we presented it in the context of an arbitrary number of players, although we did maintain symmetry. Heterogeneity of players, of course, is a real phenomenon, and can give rise to interesting policy relevant phenomena, such as tipping. It is, therefore, a natural subject for future work to extend the formal analysis framework to asymmetric interactions, and apply it to games with heterogeneous players. Another strong assumption that we have made is that of complete information, at least from the perspective of the analyst. In future work, it will be important to allow the analyst, as well as the players, to have private information, for example about the cost of investing in security.

Acknowledgments This work was supported in part by the Climate Decision Making Center (CDMC) located in the Department of Engineering and Public Policy (Cooperative Agreement between the NSF (SES-0345798) and Carnegie Mellon University), CREATE (National Center for Risk and Economic Analysis of Terrorism Events, funded by the U.S. Department of Homeland Security, award number 2007-ST-061-000001), the Center for Research on Environmental Decisions (CRED; NSF Cooperative Agreement SES-

0345840 to Columbia University), Wharton Risk Management and Decision Processes Center, and Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

8 Appendix

8.1 Description of Strategies

8.1.1 Full Feedback

Our consideration set of strategies in the full information context is

1. *Prob(I)=0.7*: play Invest with probability 0.7 (approximately the probability of Invest in early rounds of the human subject experiments)
2. *Prob(I)=0.2*: play Invest with probability 0.2 (approximately the probability of Invest in later rounds of the human subject experiments)
3. *AlwaysInvest*: Invest no matter what the opponent does
4. *NeverInvest*: Don't Invest no matter what the opponent does
5. *TFT*: classic Tit-for-Tat strategy
6. *InvestAfterLoss*: Invest after experiencing a loss
7. *InvestNAfterLoss*: A player using *InvestNAfterLoss* does *not* invest on the first round, and continues to not invest, except for the N rounds immediately following a loss, whether direct or indirect. N is set to 3 for these experiments.
8. *DontInvestAfterLoss*: Don't Invest after experiencing a loss
9. *1TitFor2Tats*: same as Tit-for-Tat except wait until the counterpart plays Don't Invest for two rounds in a row before responding with Don't Invest
10. *2TitsFor1Tat*: same as Tit-for-Tat except respond with two consecutive rounds of Don't Invest to any Don't Invest decision by the counterpart
11. *FictiousPlay*: plays a best response to the observed (empirical) mixed strategy of the counterpart

8.1.2 Partial Feedback

The set of policies used in partial feedback games is

1. *Prob(I)=0.7*: same as above
2. *Prob(I)=0.2*: same as above
3. *AlwaysInvest*: same as above
4. *NeverInvest*: same as above
5. *InvestAfterLoss*: same as above
6. *InvestNAfterLoss*: same as above
7. *DontInvestAfterLoss*: same as above
8. *TitForTatPlusLossInvest*: partial feedback analog of Tit-for-Tat, where a player responds only when the Don't Invest decision by the opponent is inferred (i.e., when he experiences the indirect loss); in addition, Invest after experiencing a loss

9. *TitForTatPlusLossNotInvest*: partial feedback analog of Tit-for-Tat, where a player responds only when the Don't Invest decision by the opponent is inferred (i.e., when he experiences the indirect loss); in addition, Don't Invest after experiencing a loss
10. *TitForTatPlusSticky*: Under full feedback a player knows whether the counterpart has invested in security during the previous rounds of play. In this strategy, the player plays a tempered form of Tit-for-Tat. The player cooperates until the the Don't Invest decision by the opponent is inferred (i.e., when he experiences the indirect loss), then defects and continues to defect until the counterpart has cooperated $N = 3$ times in a row.

References

Axelrod, R. (1980a). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(1), 3–25.

Axelrod, R. (1980b). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(3), 379–403.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books Inc.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Bartholdi, J. J., Butler, C. A., & Trick, M. A. (1986). More on evolution of cooperation. *Journal of Conflict Resolution*, *30*(1), 129–140.

Bendor, J. (1993). Uncertainty and the evolution of cooperation. *Journal of Conflict Resolution, 37*(4):709–734. doi:10.1177/0022002793037004007. URL http://jcr.sagepub.com/cgi/content/abstract/37/4/709.

Bendor, J., Kramer, R. M., Stout S. (1991). When in doubt... *Journal of Conflict Resolution, 35*(4):691–719. doi:10.1177/0022002791035004007. URL http://jcr.sagepub.com/cgi/content/abstract/35/4/691.

Bereby-Meyer, Y., Roth, A. E. (2006). The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *The American Economic Review, 96*(4), 1029–1042. ISSN 00028282. URL http://www.jstor.org/stable/30034329.

Bowling, M. (2004). *Convergence and no-regret in multiagent learning*. In Neural Information Processing Systems.

Bowling, M., & Veloso, M. (2002). Multiagent learning using a variable learning rate. *Artificial Intelligence*, *136*, 215–250.

Camerer, C., & Kunreuther, H. (1989). Decision processes for low probability events: Policy implications. *Journal of Policy Analysis and Management*, *8*(4), 565–592.

Friedman, D. (1991). Evolutionary games in economics. *Econometrica*, *59*(3), 637–666.

Fudenberg, D., & Maskin, E. (1990). Evolution and cooperation in noisy repeated games. *The American Economic Review*, *80*(2), 274–279.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: The MIT Press.

Fudenberg, D., Rand, D. G., Dreber, A. (2010). *Slow to anger and fast to forgive: Cooperation in an uncertain world*. Working paper, Harvard University, 26 May 2010.

Greenwald, A., Hall, K. (2003). *Correlated-q learning*. In International Conference on Machine Learning, pp. 242–249

Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Heal, G., & Kunreuther, H. (2005a). You only die once: Interdependent security in an uncertain world. In H. W. Richardson, P. Gordon, & J. E. Moore II (Eds.), *The economic impact of terrorist attacks* (pp. 35–56). Cheltenham: Edward Elgar.

Heal, G., & Kunreuther, H. (2005b). IDS models of airline security. *Journal of Conflict Resolution*, *49*(2), 201–217.

Kimbrough, S. O. (2012). *Agents, games, and evolution: Strategies at work and play*. Boca Raton, FL: CRC Press.

Kunreuther, H., & Heal, G. (2003). Interdependent security. *The Journal of Risk and Uncertainty*, *26*(2/3), 231–249.

Kunreuther, H., Silvasi, G., Bradlow, E., & Small, D. (2009). Bayesian analysis of deterministic and stochastic prisoner's dilemma games. *Judgment and Decision Making*, *4*(5), 363–384.

Littman,M. (1994). *Markov games as a framework for multi-agent reinforcement learning*. In International Conference on Machine Learning, pp. 157–163.

Marinoff, L. (1992). Maximizing expected utilities in the prisoner's dilemma. *Journal of Conflict Resolution, 36*(1):183–216. doi:10.1177/0022002792036001007. URL http://jcr.sagepub.com/cgi/content/abstract/36/1/183.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*, 56–58.

Ostrom, E. (2009). *Beyond markets and states: Polycentric governance of complex economic systems*. In Nobel Prize Lecture, pp. 408–444.

Pelc, A., Pelc, K. J. (2009). Same game, new tricks. *Journal of Conflict Resolution, 53*(5):774–793. doi:10.1177/0022002709339045. URL http://jcr.sagepub.com/cgi/content/abstract/53/5/774.

Powers, R., Shoham, Y. (2005). *New criteria and a new algorithm for learning in multi-agent systems*. In Neural Information Processing Systems, pp. 1089–1096.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., et al. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, *328*, 208–213.

Rogers, A., Dash, R. K., Ramchurn, S. D., Vytellngum, P., & Jennings, N. R. (2007). Coordinating team players within a noisy Iterated Prisoner's Dilemma tournament. *Theoretical Computer Science*, *377*, 243–259.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Florence, MA: Free Press.

Roughgarden, J. (2009). *The genial gene: Deconstructing Darwinian selfishness*. Berkeley, CA: The University of California Press.

Skyrms, B. (2010). *Signals: Evolution, learning & information*. Oxford: Oxford University Press.

Thomas, E. A. C., & Feldman, M. W. (1988). Behavior-dependent contexts for repeated plays of the Prisoner's Dilemma. *Journal of Conflict Resolution*, *32*(4), 699–726.

Walsh, W. E., Das, R., Tesauro, G., Kephart, J. O. (2002). *Analyzing complex strategic interactions in multi-agent systems*. In Fourth Workshop on Game Theoretic and Decision Theoretic Agents.

Wu, J., Axelrod, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution, 39*(1):183–189. doi:10.1177/0022002795039001008. URL http://jcr.sagepub.com/cgi/content/abstract/39/1/183.

Xiao, E., Kunreuther, H. (2010). *Punishment and cooperation in stochastic social dilemmas*. Working paper, Wharton School, University of Pennsylvania, 14 July 2010.

Zawadzki, E., Lipson, A., Leyton-Brown, K. (2008). *Empirically evaluating multiagent learning algorithms*. Working paper.