CENTRE FOR VOTING AND PARTIES
DEPARTMENT OF POLITICAL SCIENCE
UNIVERSITY OF COPENHAGEN

EXIT, COLLECTIVE ACTION AND POLYCENTRIC POLITICAL SYSTEMS

Peter Kurrild-Klitgaard
Professor

Centre for Voting and Parties
Department of Political Science
University of Copenhagen

Forthcoming in Special Issue of "Public Choice" in Honour of Elinor Ostrom, Nobel Laureate

CVAP Working Paper Series
CVAP WP 1/2010
ISBN: 987-87-7393-571-2
www.cvap.polsci.ku.dk

About the CVAP Working Paper Series

The CVAP Working Paper Series contains finished drafts and preliminary versions of articles and chapters within the field of studies of voting, elections, political parties, parliaments, etc., that are on their way in the publication process.

About the Centre for Voting and Parties

The Centre for Voting and Parties (*Center for Valg og Partier*, CVAP) is a research centre attached to the Department of Political Science at the University of Copenhagen. The Centre was established in 2008 and concerns itself with research within the fields of voting, elections, political parties, parliaments, etc., both at a national and a comparative level.

For more information on the Series or CVAP, visit www.cvap.polsci.ku.dk.

Exit, collective action and polycentric political systems

PETER KURRILD-KLITGAARD

Professor, Dept. of Political Science, University of Copenhagen

Abstract Elinor Ostrom and the Bloomington School's important contributions include the development of the concept of "polycentric" political systems and the demonstration that solutions to common-pool resource problems may be solved voluntarily by rational individuals, even in situations that resemble Prisoners' Dilemmas. The program, however, pays little attention to how individuals' ability to exit may affect the interaction in Prisoners' Dilemma-like situations, for worse or better. We argue why this is a worthwhile consideration and survey results from public choice and game theory.

Keywords Collective action; game theory; Prisoners' Dilemma; simulations; experiments; governance; Elinor Ostrom

1. Introduction

The award of the 2009 Nobel Prize in Economics to Elinor Ostrom (jointly with Oliver Williamson) for her analysis of economic governance and problems of the commons was a much deserved recognition of her important contributions to the study of collective action. However, it was also yet another *cadeau* to the public choice tradition, for while having its own distinct trademarks the research program of Elinor Ostrom is closely related to the public choice tradition broadly conceived.[1]

Yet at the same time, there are some interesting differences in tone between Bloomington public choice and Virginia public choice, at least in some of the works of the Buchanan-Tullock tradition. The latter often seems Hobbesian in its approach to collective action problems: Social interaction is seen as potentially characterized by severe collective action problems and government is posited as a potential, centralized and top down "Leviathan"-style solution to such problems (although also with its own set of problems) (Tullock 1974; Buchanan [1975] 1999; cf. Wagner 2008). In contrast, the Bloomington tradition—and especially Elinor Ostrom's work on governance of the commons and the spontaneous, trust-based evolution of such—could be seen as owing much more to Locke,

[1] See, e.g., Ostrom and Ostrom 1971; Ostrom [1971] 1978; Ostrom 1986b; Ostrom 1991; Ostrom and Ostrom 2004; cf. also Mitchell 1988; Aligica and Boettke 2009: 116ff, 124-36, 143ff. One of Elinor Ostrom's very first academic articles was a critical appreciation in this journal of Buchanan and Tullock's constitutional economics (Ostrom 1968).

Montesquieu, Hume, Smith, Madison, Hamilton and Tocqueville.[2] Here the focus is more on how collective action problems may be overcome locally, competitively and by the participants themselves than on how they must be solved through centralized action.

The purpose of the present paper is to highlight one aspect of this difference, while simultaneously suggesting a further perspective that may be seen as often under-articulated in both the Virginia and the Bloomington analyses: The potential importance of "exit" options in collective action settings when it comes to inducing coordination of behavior in a mutually beneficial direction. In the following we shall, first, briefly introduce the Ostroms' analyses of polycentricity and common-pool resource problems (section 2), then argue why a focus on exit-options is missing but worthwhile (section 3), and briefly survey some relevant recent results (section 4).

2. Ostrom, polycentricity and collective action

Polycentric systems

A central element and a novel concept in the work of the Ostroms has been that of "polycentricity", which grew out of academic and political debates in the 1960s and 1970s relating to how to organize local governments in the US (Aligica and Boettke 2009: 5ff). At one end of the spectrum was the view that there in each major urban area should be only one government unit—a single integrated and hierarchical system of administration, and something resembling an abolition of the separation of powers in the local government. Many reformers saw such centralized and uniform government as the solution to the supposed chaos and inefficiency created by many units and several levels. At the other end of the spectrum were some, among whom the Bloomington scholars, who argued that a centralization of the governance units cannot be the one and only universal solution: There simply are too many, too diverse circumstances and problems at local levels for one solution to fit all, and accordingly there is a need for diverse patterns of association, whereby local problems may be solved in the way that is best fitted to the individual resources, time horizons, etc. Some public goods and services might be more efficiently produced on a larger scale, while others might be produced more efficiently on a smaller. Their normative point thus was that self-government and decentralization should be allowed at the socially optimal level of governance:

[2] See, e.g., Ostrom 1990: 216; Ostrom 1992; Ostrom 1997; Aligica and Boettke 2009: 139ff, 159.

> "Overlapping service areas and duplicate facilities are necessary conditions for the maintenance of competition in a market economy. Can we expect similar forces to operate in a public economy? If we can, relationships among the governmental units, public agencies, and private businesses functioning in a public economy can be coordinated through patterns of interorganizational arrangements. Interorganizational arrangements, in that case, would manifest market-like characteristics and display both efficiency-inducing and error-correcting behavior. Coordination in the public sector need not, in those circumstances, rely exclusively upon bureaucratic command structures controlled by chief executives. Instead, the structure of interorganizational arrangements may create important economic opportunities and evoke self-regulating tendencies." (Ostrom and Ostrom 1965: 3)

For such an organization to function efficiently it will be necessary to structure the governance units so as to avoid that individuals pay for public goods and services that they do not receive, or that they receive public goods and services for which they do not pay, i.e., units must be fitted to the most efficient scale of production (cf. Aligica and Boettke 2009: 13ff).

From this insight, so to speak, sprang two parts of the subsequent Ostrom program: First, an interest in how groups and organizations in practice structure collective action most efficiently and, second, a focus upon the ideal types of "monocentrism" versus "polycentrism" and how these may support optimal outcomes in different ways. The former refers to a political system, where one and only one center of power and authority exists, with a monopoly of all the legal use of coercion, and covering the relevant geographical jurisdiction, and from which a hierarchy descends. In contrast,

> "By conceptualizing metropolitan areas as polycentric political systems, we were suggesting that a system of ordered relationships underlies the fragmentation of authority and overlapping jurisdictions that had frequently been identified as 'chaotic' and as the principal source of institutional failure in the government of metropolitan areas. We identified a polycentric political system as having many centers of decision making that were formally independent of each other." (Ostrom [1972] 1999: 52)

For the Ostroms, polycentricity is thus both an analytical concept that describes a certain feature that may exist in a more or less dominant form in the economic and political processes and the legal

and constitutional systems of the real world, and a claim about certain attractive properties of such arrangements—specifically that polycentric systems are more likely to have the incentives that will lead to self-organized, self-corrective institutional change:

> "While all institutions are subject to takeover by opportunistic individuals and to the potential for perverse dynamics, a political system that has multiple centers of power at differing scales provides more opportunity for citizens and their officials to innovate and to intervene so as to correct maldistributions of authority and outcomes. Thus, polycentric systems are more likely than monocentric systems to provide incentives leading to self-organized, self-corrective institutional change." (Ostrom 1997: n.p.)

This is so because market-like mechanisms may be able to develop competitive pressures that will tend to generate higher efficiency than may be realized by units organized through monopolies and hierarchies. This market analogy necessarily raises the question of what the positive and negative consequences may be of letting citizens, or firms or institutions, "vote with their feet" within the political system, when, say, alternatives are offered by multiple producers within the same geographical area:

> "The presence of more than a single producer of urban public goods within a metropolitan area may enable citizens to make more effective choices about the mix of services they prefer to receive than reliance upon voting mechanisms and a single producer. Multiple governments existing within a metropolitan area enable citizens to 'vote with their feet.' The presence of multiple producers within one metropolitan area may also reduce the cost for citizens of comparing the levels of output provided by different jurisdictions. Public officials who are representing one constituency in a bargaining process with other public officials over cooperative arrangements (such as contracting for services to be performed) may be able to bargain more effectively if alternative public producers are present in the area. However, it is also possible that multiple producers of some urban public goods may nullify each other's actions and lead to a reduction in net output of urban public goods. The political economist will consider the effect of competition among public agencies as an empirical question. The effect may be positive or negative depending upon the type of urban public good being considered." (Ostrom 1972: 484f)

This perspective led the Ostroms to formulate a specific proposition (as one of many) of their political economy approach: "Multiple jurisdictions with different scopes and scales of organization allow citizens to make better effective choices when selecting packages of services most important to them, to better articulate their preferences and concerns, and, if necessary, to move to other jurisdictions." (Ostrom 1972: 478f). The affinity with Tiebout's work on mobility should be clear (Tiebout 1956)—and actually is no accident, given that Vincent Ostrom and Tiebout were early collaborators on analyses of the effects of mobility for public decision-making and public finance (Ostrom, Tiebout and Warren 1961; cf. also Hirschman 1970). For the present purposes it is particularly important how the Ostroms have repeatedly noted that the theoretical and empirical possibility of opting in or out of interactions potentially must be very important for the understanding of the dynamics of polycentric arrangements (cf. Aligica and Boettke 2009: 46ff).

Collective action and the Prisoners' Dilemma

From the Ostroms' interest in political economy, social organization and polycentricity has grown an impressive empirical research agenda. The canonical way of representing social dilemmas and collective action problems has for decades been the notorious Prisoners' Dilemma, which Elinor Ostrom herself frequently has used as the analytical point of departure. Following her and many others,[3] the game can be defined as a situation of interaction between two actors, I and II, who are faced with the one-shot choice of either cooperating (*C*) or defecting (*D*) from a collective action, and thus having four possible outcomes. If both cooperate they both reap the cooperator's reward (*r*), while if they both defect they will both endure the defector's punishment (*p*). If one player gives in to the temptation to defect, while the other cooperates, the former will receive a traitor's payoff (*t*) while the latter gets the sucker's payoff (*s*). This general form of the payoffs of what can be called a Collective Action Game is given in Figure 1.

[3] See, e.g., Ostrom 1990: 3ff; 39ff; Ostrom 2000: 137ff; Ahn et al. 2001; Ahn, Ostrom and Walker 2003: 296ff; cf. Rapoport and Chammah [1965] 1970, who first used the particular form given here.

Figure 1. General Payoffs for a one-shot 2 x 2 Collective Action Game

		II	
		C	D
I	C	r r	t s
	D	s t	p p

Given the matrix, a Prisoners' Dilemma is usually defined by two relations: $t > r > p > s$ and $2r > t + s$. The first relation alone is considered sufficient to make the game a PD, but the second is often added so as to make mutual cooperation better than coordinated alternation of cooperation (cf. Rapoport and Chammah [1965] 1970: 34; Schuessler 1989: 746; Boone and Macy 1999: 32; Axelrod 2000: 131; Janssen 2008). If the size of the payoffs for each player indeed are such that the preference ordering over the four individual payoffs is $t > r > p > s$, then we have the familiar logic that the dominant strategy for both players is to defect, because no matter what the other player does, playing *D* will earn himself a more highly valued outcome. This will result in the unique Nash equilibrium (*D*, *D*), despite the fact that this is Pareto inferior to the one where both cooperate (*C*, *C*).

As is well-known this logic can be applied to a wide variety of social situations. In fact, by changing the values of the payoffs (thereby changing the individual preference orderings) this framework allows for a potentially infinite multitude of games portraying various situations, including such fundamental ones as exchanges or contributions to a common project. If, for example, in contrast to the PD, it is assumed that, say, a Hobbesian "Leviathan" imposes costs on Defection and that these are of such a size, that the total payoff from *Defection* becomes smaller than the total payoff from *Cooperation*, then we get preference orderings such as, e.g., $r > s > t > p$ and quite different outcomes.

PD-games have, in fact, often been the point of departure of Elinor Ostrom herself, while she simultaneously, in effect, has devoted a large portion of her research for the uncovering of how real-world collective action situations do *not* resemble such (Ostrom 1990: 46f; 217f). Quite often individuals are able to cooperate, and when the situation stops resembling the one-shot 2 x 2 PD, and involves possibilities for, e.g., repeated interaction, sanctions, compensations, communication,

changes in the number of players, the amount of cooperation (e.g. size of donations), asymmetries in preferences, etc., a number of theoretical possibilities occur, just as data, from computer simulations, laboratory experiments and the more complex real world, shows that cooperation may become pervasive (cf., e.g., Ostrom 2000).

If this is true—that social phenomena should be seen as the outcomes of the interaction of rational individuals, and if the equivocation of social dilemmas and the Prisoners' Dilemma does not square off well with empirical observations—then the subject must be approached differently, and hence springs Elinor Ostrom's original, extensive and varied contributions to the theoretical and empirical study of how solutions to common-pool resource problems take place.

Realism and exit

Critics of the applications of the PD have often suggested that it is not "realistic". Obviously, this is to a large extent beside the point, since no stylized representation of interaction will ever be able to capture all aspects of real world interaction between real human beings, and the important question rather is whether the game captures "enough" of the essentials of naturally occurring situations to be a relevant metaphor.

This is no doubt the case in many interactions. But it should be equally clear that there are important aspects of the interaction of individuals, say in the marketplace or civil society more broadly, where the PD may not be adequate. In fact, of the billions and billions of daily interactions in the market place and in civil society, the vast majority consist of more complicated choices and less clear outcomes, and not just because the players are afraid of being detected and prosecuted by a Hobbesian Leviathan. People who meet in the marketplace, most of the time do not have a simple choice between, in effect, stealing or exchanging—they can (at least in developed markets) more often take their business elsewhere, if, say, the other market participant has a low credit rating, a bad image, etc. Similarly with citizens trying to organize a joint-venture of a common-pool resource type or participate in an organization or some form of alliance: Their choice often is not simply whether to contribute to the project or free-ride, but may include an option of simply not taking part and perhaps instead take up the task with someone else (cf. Orbell, Schwartz-Shea and Simmons 1984: 148; Tullock 1985: 1073f). It all depends, as the Ostroms are crucially aware, of the specifics of the "action situation" and the rules of the "action arena" (cf. Ostrom 1986a).

Elinor Ostrom has quite convincingly argued that the Prisoners' Dilemma is not the best way to model many, perhaps most, social interactions in collective action situations. Yet, she does not herself explicitly model her common resources problems as including exit options (cf., e.g., the games in Ostrom 1990: 4ff; 10ff; 15ff), and that despite the fact that she is quite familiar with some of the attempts to model exit (cf. Ostrom 1998: 12). Rather, she often begins from the point where the individuals as a group have to decide whether to cooperate or defect within, so to speak, the framework of collective choice and even suggests that, e.g., business exchanges (where exit options certainly are present) may have the structure of a PD (cf. Ostrom 1990: 15f).

That is perhaps understandable: Doing so might lead to considerably more complex models and thus conflict with an ambition to keep things (relatively) simple. Second, the possibility of exit, especially when it is cheap, might be steering outcomes in a completely different direction: Exit could easily be seen as undermining solutions to collective action problems and potentially lead to larger conflicts. Ostrom has stated in general terms relating to the closely related topic of migration:

> "All economic and political organizations are vulnerable to threats, and self-organized resource-governance regimes are no exception. Both exogenous and endogenous factors challenge their long-term viability. … Major migration (out or into an area) is always a threat [to the strength of the norms of trust and reciprocity] that may or may not be countered effectively. Out-migration may change the economic viability of a regime due to loss of those who contribute needed resources." (Ostrom 2000: 153).

A third reason is that Ostrom perhaps has been more interested in how already existing cooperatives with high costs of exit might function, rather than those instances where individuals can easily exit, or where they ex ante need to make a decision on whether or not to enter at all.

On the other hand, there are reasons why exit should be considered and why it may seem strange that it is not considered by Elinor Ostrom. First of all, as indicated, exit is thematically closely related to the notion of polycentricity found almost everywhere, implicitly or explicitly, in the Ostroms' works on subjects such as political theory, federalism, public economics, governance and indeed common-pool resources. A prime element in this is that there are multiple authorities with overlapping jurisdictions, which may operate in ways reminiscent of markets, but if that is the case, then it almost begs the question of whether it is possible for individuals to opt out of (or into) one or more institutional arrangements. Furthermore, in several of the real-world common

resource situations Ostrom study, there would seem to be exit options that might be considered, no matter whether the groups are lawyers or farmers (cf., e.g., Ostrom 1990: 25, 61-65, 69-88). This may be important: Peasants who cannot sell their farms and cannot opt out of communal farming arrangement may behave quite differently with regard to how much they are willing to suffer, then peasants who may be able to dispose freely of their property or who may opt in or out of the communal enterprise.

3. Adding exit-options to social dilemmas

In the following we shall try to briefly highlight the points just made. In order to do so, we will engage in a relatively simple but nonetheless crucial reformulation of the Prisoners' Dilemma, where we add a bit of realism and at the same time highlights the importance of exit as an option. This will be easy, as there is a small but rapidly growing literature by public choice economists and game theorists arguing that it may be exactly the possibility of exit from future interactions that may sustain trust, develop social capital and support endogenous, non-centralized cooperation and perhaps even be a check on tyranny (e.g., Orbell, Schwartz-Shea and Simmons 1984; Tullock 1985; Schuessler 1989; Vanberg and Congleton 1992; Orbell and Dawes 1993; Yamagishi and Hayashi 1996; Macy and Skvoretz 1998; Boone and Macy 1999; Edk-Group 2000; Congleton and Vanberg 2001; Kurrild-Klitgaard 2002; Phelan, Arend and Seale 2005; Arend and Seale 2005; Seale, Arend and Phelan 2006; Janssen 2008; Arend 2009; Ahn, Esarey and Scholz 2009).[4]

The two-person one-shot PD with Exit game

Let us assume that the two players of the standard PD have an additional possibility of choosing to "exit" a possible interaction, where someone simply refuses to deal with another person (the action labeled *E*). If we add these changes to the Collective Action Game of Figure 1, we obtain a new type of game, as shown in the 2 x 3 matrix of Figure 2.

[4] Robert Axelrod has singled out the research on exit as one of the most important advances in the study of cooperation in recent decades (Axelrod 2000). Some of the points also connect more generally to discussions of the PD as a model of anarchic interactions; cf., e.g., Taylor 1987; Powell and Stringham 2009.

Figure 2. General Payoffs for a one-shot 2 x 3 State-of-Nature-with-Exit Game

		II		
		C	D	E
I	C	r r	t s	n n
	D	s t	p p	n n
	E	n n	n n	n n

In order to make this new game comparable with the standard 2 x 2 collective action games we add the payoff of "no payoff" (*n*) to the familiar ones of *p*, *r*, *s* and *t*. It is obvious that the game's character and outcome will depend upon the relative values of the payoffs. If we make the by now standard assumption in this line of research, that $n = 0$ and that *t* and *r* have a positive value while *p* and *s* have a negative, then the ranking of the payoffs are $t > r > n > p > s$, and so the basic structure of the game's four most north-western cells will be that of the Prisoners' Dilemma albeit with the now added possibility of Exit. This is, in other words, what we might call the Prisoners' Dilemma-with-Exit Game (PDE).

If analyzed as a one-shot simultaneous move game, then instead of having *Defect* as the strictly dominant strategy and the outcome (*D*, *D*) as the one and only Nash equilibrium, this game has another dominant strategy, *Exit*, and three pure strategy Nash equilibria in the form of the outcomes (*D*, *E*), (*E*, *D*) and (*E*, *E*), i.e., where the exit option is used by at least one player. This is so, because if one player expects that the other player will *Cooperate*, the best option is to *Defect*, but since the game is symmetrical each player reasons similarly, and then both will want to *Defect*—in which case their best option is to *Exit*, and the Nash equilibrium is mutual exit (cf. Arend and Seale 2005: 1061f; Janssen 2008: 460). Similarly, if the game is solved as an extensive form game with sequential moves, the subgame perfect equilibrium will be one where at least one player will choose *Exit* (cf. Phelan, Arend and Seale 2005: 343; Seale, Arend and Phelan 2006: 62f). Now, an outcome of, say (*E*, *E*) certainly is Pareto inferior to (*C*, *C*), but while not Pareto optimal it is, by essentially being the status quo, for the very same reason Pareto superior to (*D*, *D*).

Iterated two-person PDE-games

The one-shot PDE game suggests that if players can exit from something otherwise resembling a PD, then they may be able to avoid the unattractive outcome of mutual defection, albeit without realizing the socially optimal outcome. But how about repeated interactions? Although the one-shot game merely secures the less than best result that it is better to stick with the status quo than be a sucker, this may change significantly, if the interaction is repeated. Specifically, rational individuals will now know that they may face the same players with some probability in the future—who may make their choice of future behavior dependent on the other players' past behavior, and hence punish them if they behave in a non-cooperative way. This result has been well-known for decades when it comes to repeated plays of the PD (cf. Axelrod 1984; Taylor 1987), and quite obviously similar results may obtain for repeated plays of the PDE. This, of course, also includes the result that if the number of interactions is finite and known, then it will be rational to defect in the last round, and this will then "spill back" all the way to the first round—in which case *Exit* will be chosen (Schuessler 1989; Seale, Arend and Phelan 2006).

But can we say something more specific? Two very different perspectives open up, if exit is possible and interactions are repeated (cf. Schuessler 1989: 737; Boone and Macy 1999: 32ff; Arend and Seale 2005: 1058; Seale, Arend and Phelan 2006: 61f). On one hand, the possibility of exit may lessen cooperation by reducing the extent to which players perceive themselves to be dependent on others—e.g., the ability to punish others who are defecting. If they can exit future interactions, then the mobility and anonymity of players may mean that it becomes rational for individuals to engage in "hit and run" behavior. To give a metaphorical example: The behavior of a cab driver in a small town, where he knows all citizens and can expect to have them as his customers perhaps regularly in the future, may be quite different from a cabbie at, say, New York City's JFK airport. Lack of selectivity and presence of mutual dependence may induce cooperation.

On the other hand, the logic of exit may also work exactly the other way, so as to support cooperative behavior. If players can remember past interactions and freely choose whom they are willing to play with, then players may ignore the untrustworthy and seek out the trustworthy. Players cannot afford to misbehave, and this may lead to an increase in cooperation relative to games without such exit options. Again, to put it metaphorically, a customer may be treated quite differently by, say, a limousine driver in a competitive cab market than he may by a company having a government monopoly on all transportation.

4. A short survey of variations in the PD with Exit game

The previous theoretical expectations are testable, and a number of studies have indeed tried to go beyond the non-iterated interactions.

Rudolf Schuessler has conducted a series of computer simulations involving a large number of repeated interactions in something resembling two-person PDEs. The first of these (Schuessler 1989) involves strategy contests, where strategies have no memory, effects of reputation do not exist, and defectors cannot be identified, traced or held responsible. In other words, the conditions for establishing cooperation are the worst possible. Nonetheless, the results are that cooperation evolves. Specifically, a strategy called CONCO ("conditional cooperation") begins by playing *Cooperate*, never plays *Defect* first, but plays *Exit* as soon as being exploited by a defector. Pitted against a number of defecting strategies, CONCO not only survives but also wins a contest, although with the strategy ALL D ("always *Defect*") as a close competitor, depending on the relative size of the payoffs and the number of iterations. Cooperation is maintained even if a moderate probability of the breaking apart of mutually beneficial interactions is added.

The second study, conducted by Schuessler with a number of collaborators (Edk-Group 2000), is basically identical but includes a wider set of strategies (15) and allows the players to keep interacting with defectors or leave them. In other words, one misstep by a player does not lead his partner to exit all future interactions but may give him a second (or further) chance, just as the players are given more information about the past behavior and intentions of the other players. The strategies included not only CONCO but also Axelrod's famous TIT-FOR-TAT strategy, where a player first cooperates and then in next round mimics the other player's behavior in the previous round (Axelrod 1984), as well as a number of shrewd, opportunistic strategies such as MALLD (*Defect* and *Exit*, i.e., "hit-and-run"). The clearly best performing strategy is the very robust CONCO, although the second best is a rather wicked opportunist strategy that adapts and defects. TIT-FOR-TAT performs with some mediocrity of the strategies considered, although far from worst. Perhaps most importantly, Schuessler's results show that hit-and-run exploitation is a very bad strategy, although some forms of clever opportunism pays fairly well.

Roger Congleton and Viktor Vanberg conducted two series of simulated round-robin tournaments with repeated interactions between rival strategies for how to play a PDE game. In an early version the simulations operated with two player contests (Vanberg and Congleton 1992), while a later used a three-person setting (Congleton and Vanberg 2001). In the two-player simulations were included such classic strategies as TIT-FOR-TAT, COOPERATOR (always

Cooperate), FREE-RIDER (always *Defect*), as well as HIT-AND-RUN (always *Defect* and, in this formulation, *Exit* permanently if the current payoff is negative, i.e., reminiscent of Schuessler's MALLD), as well as a number of adaptive strategies that condition their own behavior on the observed behavior of the others. They found that a strategy of so-called PRUDENT MORALIST may do quite well. Such a strategy is basically identical to Schuessler et al.'s CONCO (above), i.e., a player initially always cooperates, but refuses to play with players who have defected in the past. In repeated plays of PDE, a PRUDENT MORALIST will beat a number of other attractive strategies, including not only a NAÏVE MORALIST (who always plays and always cooperates) and an OPPORTUNIST (who always plays and always defects), but also TIT-FOR-TAT. A second version of the tournament demonstrated that in repeated PDE games with more than a couple of interactions a PRUDENT MORALIST has the highest total payoff and his presence in the games increase the payoffs of the other players.

The high levels of cooperation found in tournaments based on two-person games could be a function of the two-player restriction (cf. Axelrod 1984: 11), and so Congleton and Vanberg replicated the analysis but as a three-player tournament and with a total of nine strategies—most of them the same used for the two-player simulations but modified for three players and also including an additional retributive behavioral program that only punishes free-riders, albeit at a personal cost. Many of the results were surprisingly similar to those of the two-player simulations: In the simplest form of the tournament the cooperative strategies (COOPERATOR, TIT-FOR-TAT, PRUDENT MORALIST) do better and better over time, while the defection oriented strategies (FREE-RIDER, HIT-AND-RUN) have declining results as the number of interactions increase. The strategy with the highest average pay-off over all rounds is PRUDENT MORALIST, i.e., the one that begins by cooperating but exits when current payoffs become negative. The second best performing strategy, however, is it's mirror, HIT-AND-RUN, i.e., the strategy which first defects and then exits when the current payoff becomes negative. In general, the results showed that the possibility of exit allows individuals to exit from dysfunctional groups; this increases the potential benefits of cooperation and reduces those of free-riding. Especially long-term interactions indicate that the sorting that exit options allow is sufficient to generate a larger number of interactions by conditional cooperators.

A more recent set of studies also simulate round-robin tournament of competing strategies in PDEs (Phelan, Arend and Seale 2005). A total of 12 strategies were compared, and the three best performing strategies were very close to each other and were all different types of conditional strategies using some form of TIT-FOR-TAT, but where *Exit* is also used if triggered, i.e., if things

get too unattractive. For example, the best performing strategy, TIT-FOR-TAT OR EXIT, initially plays the familiar tit for tat strategy but plays *Exit* after two instances of *Defect* when the opportunity costs are high; when they are low, it plays *Exit* after three times *Defect* or when average winnings fall below opportunity costs. The classic TIT-FOR-TAT strategy came in sixth in the tournament, while a strategy called QUICK EXIT, which cooperates on every move until it is met with *Defect*, after which it *Exits* (i.e., similar to CONCO and PRUDENT MORALIST above), came in fourth. The worst performing strategies were some that all shared the featured of not being "nice", i.e., of not beginning by cooperating and by generally being opportunistic to some degree. Phelan et al. also used Monte Carlo simulations to test variations of the winning strategy in the hope of developing a more robust solution. Such were not found, and the over-all conclusion was that the classic TIT-FOR-TAT strategy is not the best when games have exit options; these are rather strategies that are willing to use exit when "enough is enough".

In a number of simulation experiments by Yamagishi and Hayashi conducted evolutionary tournaments similar to Axelrod's (Axelrod 1984), but with added exit options (e.g., Yamagishi and Hayashi 1996). What they termed the OUT-FOR-TAT was basically similar to the above strategies of CONCO, PRUDENT MORALIST and QUICK EXIT: The player first cooperates; if the other player also cooperates, OUT-FOR-TAT then proceeds to reciprocate, as in TIT-FOR-TAT. But if the other player defects this strategy does not also play *Defect* (as TIT-FOR-TAT) but rather exits. When OUT-FOR-TAT finds a cooperative partner, he remains with that partner while avoiding defectors. The strategy was overall the most effective strategy, showed itself highly robust and outperformed TIT-FOR-TAT.

One study points in a different direction, namely Boone and Macy who did not find any significant differences in behavior when subjects were allowed to exit (Boone and Macy 1999). Other studies based in computer simulations of repeated PDEs have instead of the random matching of players of Schuessler, Congleton and Vanberg et al. used preferential partner selection, where players can chose partners, and where those chosen can refuse based on the known history of interactions with other players (Stanley, Ashlock and Tesfatsion 1994; Ashlock et al. 1996; Majeski et al. 1997; Janssen 2008). In that case, cooperation becomes even more prevalent and robust. If there is also a so-called "waiting penalty" for exit, then the level of cooperation a population can sustain will depend on the size of exit penalty as well as on how partner selection is determined (Macy and Skvoretz 1998).

Several different laboratory experiments have been made in order to analyze the dynamics of interactions resembling iterated PDEs. Among the interesting results are that a theoretical

expectation that it would cooperation-inclined players who would tend to exit from interactions (out of fear of being submitted to "hit-and-run"- behavior) more frequently than defection-inclined players is not supported: In fact, experiments with groups of nine players have shown that the exact opposite tends to be the case (Orbell, Schwartz-Shea and Simmons 1984). Orbell and Dawes (1993) conducted experiments were they grouped players into two groups: Some playing regular PD-games and others playing PDE-games. On average the participants in the latter set of games came out with ca. 12 pct. higher profits at the end of the play than those in the former set. Overall, the experiments showed that when players are free to accept or reject play in PD-games, the aggregate social welfare increases. They also show the rather optimistic result, that it is intending cooperators who choose to play, while intending defectors who choose not to play, and that when free to accept or reject it the payoffs to intending cooperators tend to become larger than those of intending defectors.

Laboratory experiments have been conducted by Jin, Hayashi and Shinotsuka where PD games were played, either with or without selective play (Jin, Hayashi and Shinotsuka 1996). Players in four-person groups could rank-order their preferred partners, and those who preferred each other formed committed relationships with higher levels of cooperation than observed without selective play. As such there was widespread use of the OUT-FOR-TAT strategy (Yamagishi and Hayashi 1996). There was, however, little over-all difference between the level of cooperation under selective play and in forced interactions, because of low levels of cooperation between the de-selected partners.

In experiments conducted by Boone and Macy there were also low levels of cooperation, when an exit option was combined with a forced, random pairing of a new partner (Boone and Macy 1998). Boone and Macy conducted another set of experiments, using an on-line card game called "Trump" (Boone and Macy 1999). In one they found that defensive players cooperate more, when they have an option to exit than when paired with a permanent partner, while aggressive players cooperate less; in the other they found that defensive players cooperate more when exit is optional than when it is mandatory.

Among Richard Arend et al.'s several studies some include laboratory experiments (Seale, Arend and Phelan 2006), where the results actually were the opposite of their theoretical expectations (Arend and Seale 2005), i.e., participants exhibited more opportunistic behavior as opportunity

costs increase. Among their other results were, that knowing when the interaction will end, decreases cooperation.[5]

5. Conclusion

The previous points and the surveyed results from more than two decades of studying theoretical and empirical variations of Prisoners' Dilemmas with exit-options highlight three important perspectives that are fully compatible with the Ostrom program, and which may might have important consequences for the analysis of collective action problems.

First, collective action dilemmas resembling PD may exist, but frequently it may be more relevant to model the interaction as PDEs. This is especially the case if it is possible for individuals to exit. Second, if settings that superficially look like a PD-style situation the outcome need not necessarily be the detestable one of mutual *Defection*: Quite often inaction—rather than counter-action—may be a preferable outcome (individually and socially) to one of universal defection, even if it is still Pareto inferior to mutual *Cooperation*. Third, there are good reasons—theoretical and empirical—to believe that when collective interactions resemble PDEs interaction will not result in the worst of the possible outcomes, and when the interaction is iterated, conditionally cooperative strategies, i.e. those that combine being "nice" with exiting from interaction with non-cooperative players, are superior. Together such analyses of the effects of exit options suggest that in settings where exit is possible, reputation and trust will be important and indeed result in cooperative behavior by individuals and may simultaneously limit the incentives for predatory behavior. It thereby enhances our understanding of why market economies generally do not seem to have the wicked character of a Prisoners' Dilemma (Tullock 1985).

Why is this important in the present context? The Ostroms' work on polycentric social orders and on the spontaneous solution to collective action problems is both impressive and important. However, a logical next step for deepening our understanding of the nature of social interactions will be to extend similar analysis to a more general examination of "mobility", physical and virtual, for sustaining (or destroying) social interaction. If the results stand up, there may be important normative perspectives too: Adding "exit" options to institutional arrangements, where such may not exist, may be the solution to what Vincent Ostrom has called the "Faustian bargain of human

[5] For some other related work on PDEs or similar interactions, with different specifications of the number of players, the amount of information, the size of payoffs, the number of interactions, etc., see, e.g., Rasmusen and Hirshleifer 1989; Stanley, Ashlock and Smucker 1995; Hauk 2001; Hauk and Nagel 2001; Arend and Seale 2005; Ahn, Isaac and Salmon 2009; Ahn, Esarey and Scholz 2009.

societies", i.e., the ongoing threat of chaos and the ditto threat of tyranny. With no rules, we will have chaos, but by centralizing and concentrating coercive force in the hands of one institution, we are also potentially creating tyrants. But maybe those are not the two only relevant alternatives.

References

Ahn, T. K., Esarey, J., & Scholz, J. T. (2009). Reputation and cooperation in voluntary exchanges: Comparing local and central institutions. *Journal of Politics 71*(2), 398-413.

Ahn, T. K., Isaac, R. M., & Salmon, T. C. (2009). Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics 93*, 336-351.

Ahn, T. K., Ostrom, E., Schmidt, D., Shupp, R., & Walker, J. M. (2001). Cooperation in PD games: Fear, greed, and history of play. *Public Choice 106*, 137-155.

Ahn, T. K., Ostrom, E., & Walker, J. M. (2003). Heterogeneous preferences and collective action. *Public Choice 117*, 295-314.

Aligica, P. D. & Boettke, P. J. (2009). *Challenging institutional analysis and development: The Bloomington School*. Abingdon: Routledge.

Arend, R. J. (2009). Reputation for cooperation: Contingent benefits in alliance activity. *Strategic Management Journal 30*, 371-385.

Arend, R. J. & Seale, D. A. (2005). Modeling alliance activity: An iterated Prisoners Dilemma with exit option. *Strategic Management Journal 26*(11), 1057-1074.

Ashlock, D., Smucker, M. D., Stanley, E. A., & Tesfatsion, L. S. (1996). Preferential partner selection in an evolutionary study of Prisoner's Dilemma. *BioSystems 37*(99), 125.

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Axelrod, R. (2000). On six advances in cooperation theory. *Analyse & Kritik 22*, 130-151.

Boone, R. T. & Macy, M. W. (1998). Dependence and cooperation in the game of trump. *Advances in Group Processes 15*, 161-185.

Boone, R. T. & Macy, M. W. (1999). Unlocking the doors of the Prisoner's Dilemma: Dependence, selectivity and cooperation. *Social Psychology Quarterly 62*(1), 32-52.

Buchanan, J. M. ([1975] 1999). *The limits of liberty: Between anarchy and leviathan*, The Collected Works of James M. Buchanan, Vol. 7, Brennan, G., Kliemt, H., & Tollison, R. D. (Eds.). Indianapolis: Liberty Fund.

Congleton, R. D. & Vanberg, V. J. (2001). Help, harm or avoid? On the personal advantage of dispositions to cooperate and punish in multilateral PD games with exit. *Journal of Economic Behavior and Organization 44*, 145-167.

Edk-Group. (2000). Exit, anonymity and the chances of egoistical cooperation. *Analyse & Kritik 22*, 114-129.

Hauk, E. (2001). Leaving the prison: Permitting partner choice and refusal in Prisoner's Dilemma games. *Computational Economics 18*(1), 65-87.

Hauk, E. & Nagel, R. (2001). Choice of partners in multiple two-person Prisoner's Dilemma games. *Journal of Conflict Resolution 45*(6), 770-793.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Cambridge: Harvard University Press.

Janssen, M. A. (2008). Evolution of cooperation in a one-shot Prisoner's Dilemma based on recognition of trustworthy and untrustworthy agents. *Journal of Economic Behavior and Organization 65*, 458-471.

Jin, N., Hayashi, N., & Shinotsuka, H. (1996). An experimental study of Prisoner's Dilemma network: The formation of commitment among selective dyads. *Japanese Journal of Experimental Social Psychology 33*, 21-30, [published in Japanese].

Kurrild-Klitgaard, P. (2002). Opting-out: The constitutional economics of exit. *American Journal of Economics and Sociology 61*(1), 123-158.

Macy, M. W. & Skvoretz, J. (1998). The evolution of trust and cooperation among strangers. *American Sociological Review 63*, 636-660.

Majeski, S. J., Linden, G., Linden, C., & Spitzer, A. (1997). A spatial iterated Prisoners Dilemma game simulation with movement. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulation Social Phenomena* (pp. 161-167). Berlin: Springer.

Mitchell, W. C. (1988). Virginia, Rochester and Bloomington: Twenty-five years of public choice and political science. *Public Choice 56*, 101-119.

Orbell, J. M. & Dawes, R. M. (1993). Social welfare, cooperators' advantage and the option of not playing the game. *American Sociological Review 58*, 787-800.

Orbell, J. M., Schwartz-Shea, P., & Simmons, R. T. (1984). Do cooperators exit more readily than defectors? *American Political Science Review 78*, 147-162.

Ostrom, E. (1968). Some postulated effects of learning on constitutional behavior. *Public Choice 5*(Fall), 87-104.

Ostrom, E. (1972). Metropolitan reform: Propositions derived from two traditions. *Social Science Quarterly 53*, 474-493.

Ostrom, E. (1986a). A method of institutional analysis. In F. X. Kaufmann, G. Majone, & V. Ostrom (Eds.), *Guidance, Control and Evaluation in the Public Sector* (pp. 459-475). Berlin & New York: W. de Gruyter.

Ostrom, E. (1986b). An agenda for the study of institutions. *Public Choice 48*, 3-25.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
Ostrom, E. (1992). Community and the endogenous solution of common problems. *Journal of Theoretical Politics 4*(3), 343-351.
Ostrom, E. (1997). The comparative study of public economies. Acceptance paper for the Frank E. Seidman Distinguished Award in Political Economy, Rhodes College, Memphis, TN, September 26, 1997. http://www.indiana.edu/~workshop/publications/materials/seidmanweb.htm
Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. *American Political Science Review 92*(1), 1-22.
Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives 14*(3), 137-158.
Ostrom, E. & Ostrom, V. (2004). The quest for meaning in public choice. *American Journal of Economics and Sociology 63*(1), 105-147.
Ostrom, V. ([1971] 1978). *The political theory of a compound republic: Designing the American experiment*. 2 ed. San Francisco: ICS Press.
Ostrom, V. (1991). *The meaning of American federalism: Constituting a self-governing society*. San Francisco: ICS Press.
Ostrom, V. ([1972] 1999). Polycentricity. In M. D. McGinnis (Ed.), *Polycentricity and Local Public Economies: Readings From the Workshop in Political Theory and Policy Analysis* (pp. 52-74). Ann Arbor: University of Michigan Press.
Ostrom, V. & Ostrom, E. (1965). A behavioral approach to the study of intergovernmental relations. *Annals of the American Academy of Political and Social Science 359*, 137-146.
Ostrom, V. & Ostrom, E. (1971). Public choice: A different approach to the study of public administration. *Public Administration Review 31*(March/April), 203-216.
Ostrom, V., Tiebout, C. M., & Warren, R. (1961). The organization of government in metropolitan areas: A theoretical inquiry. *American Political Science Review 55*(4), 831-842.
Phelan, S. E., Arend, R. J., & Seale, D. A. (2005). Using an iterated Prisoner's Dilemma with exit option to study alliance behavior: Results of a tournament and simulation. *Computational and Mathematical Organization Theory 11*, 339-356.
Powell, B. & Stringham, E. P. (2009). Public choice and the economic analysis of anarchy: A survey. *Public Choice 140*(3-4), 503-538.
Rapoport, A. & Chammah, A. M. ([1965] 1970). *Prisoner's Dilemma*. 2 ed. Ann Arbor: University of Michigan Press.
Rasmusen, E. B. & Hirshleifer, D. (1989). Cooperation in a repeated Prisoner's Dilemma with ostracism. *Journal of Economic Behavior and Organization 12*, 87-106.
Schuessler, R. (1989). Exit threats and cooperation under anonymity. *Journal of Conflict Resolution 33*(4), 728-749.
Seale, D. A., Arend, R. J., & Phelan, S. E. (2006). Modeling alliance activity: Opportunity cost effects and manipulations in an iterated prisoner's dilemma with exit option. *Organizational Behavior and Human Decision Processes 100*, 60-75.
Stanley, E. A., Ashlock, D., & Smucker, M. D. (1995). Iterated Prisoner's Dilemma with choice and refusal of partners: Evolutionary results. In *Advances in Artificial Life* (pp. 490-502). Lecture Notes in Computer Science, Vol. 929, Third European Conference on Artificial Life Granada, Spain, June 4–6, 1995 Proceedings.
Stanley, E. A., Ashlock, D., & Tesfatsion, L. S. (1994). Iterated Prisoner's Dilemma with choice and refusal of partners. In C. Langton (Ed.), *Artificial Life III* (pp. 131-175). Vol. 17, Boston: Addison-Wesley.
Taylor, M. (1987). *The possibility of cooperation*. Cambridge: Cambridge University Press.
Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy 64*(October), 416-424.
Tullock, G. (1974). *The social dilemma: The economics of war and revolution*. Blacksburg: Center for Study of Public Choice.
Tullock, G. (1985). Adam Smith and the Prisoner's Dilemma. *Quarterly Journal of Economics 100*(Supplement), 1073-1081.
Vanberg, V. J. & Congleton, R. D. (1992). Rationality, morality and exit. *American Political Science Review 86*(2), 418-431.
Wagner, R. E. (2008). Finding social dilemma: West of Babel, not east of Eden. *Public Choice 135*, 55-66.
Yamagishi, T. & Hayashi, N. (1996). Selective play: Social embeddedness of social dilemmas. In W. B. G. Liebrand, D. M. Messick, & H. A. M. Wilke (Eds.), *Frontiers in Social Dilemmas Research* (pp. 363-384). Berlin: Springer.