2. The Problem of Social Order

In this section of the book, I will seek to identify elements both of cooperation and conflict in social interaction. Why is it that we observe law enforcing institutions? Is there a structural need for sanctions? Are rules generally, or are at least certain types of rules, self enforcing? And, why do individuals break those rules that they had adopted previously?

Let me commence my discussion with a brief outline of a set of concepts that share a somehow romantic vision of society and the state. According to these concepts, the fundamental problem of social order is one of finding (and promulgating) rules of behavior that, once established, will be followed voluntarily by the individuals, without means of enforcement being necessary[32].

These concepts, although sharing the same basic idea of social stability and harmony, however, differ in respect of the presumed motivations that induce the individual to comply. According to a more romantic concept, individuals follow rules inspired and motivated by their inherently benevolent nature. Other more agnostic contributions[33] concede to the self-interested nature of human beings, but assert that individuals nonetheless will conform to such rules. They will comply in view of the fact that the law benefits everyone in society, including themselves. The common good for the group (in terms of the functional role of the rules employed) will, therefore, induce individuals to obey the rules.

One can recast this approach to the problem of social order as one of coordination where a set of rules once found or agreed upon will be self stabilizing. If this is the appropriate conceptual framework, we can

[32] Such concept need not necessarily be one of natural law. The standards of behavior may be simply agreed upon by the individuals involved.

[33] This is, most prominently, the position of the functionalist school of sociology. As is to be noticed, also the traditional law and economics explanation of the law has substantial functionalist sidetastes. See also the critique by Coleman (1988).

seek to capture its implicit suppositions with respect to the structure of social interaction by game theory, and more specifically by coordination games.

2.1. Coordination Problems

2.1.1. A General Outline

Generally, game-like settings are characterized by a mutual dependency of individual choices, such that each player's payoff depends both on his own and on the other player's (players') strategy (strategies). Players, hence, can only choose strategies, they cannot choose outcomes. Outcomes emerge from the separate choices that individuals make.

Coordination problems have the distinctive property that there is a coincidence of interests among the players involved that overwhelms all sources of possible conflict[34]. Hence, only concerted choices yield mutually satisfactory outcomes: the players win and loose together. If there was only one positively valued outcome, coordination would be an easy task: each actor would choose the only promising strategy. Coordination problems, however, involve a certain element of ambiguity in respect to the other player's choices. Each player's choice, hence, depends on what he assumes his fellow will choose. The nontrivial problem consists in picking the right choice.

For a further elaboration of coordination problems, let me commence with the well-known example of a phone call being cut off, such that the two persons involved find themselves confronted with the choice either to call back or to wait. It is obvious that neither simultaneous efforts to call back nor to wait would be beneficial (in terms of the players' own evaluations). The players' interests coincide insofar as they want one of them to wait and the other to call back. Yet, each of them has to choose according to his expectations of the behavior of the other. This can be best expressed in a matrix, in which the off-diagonal cells are those wherein the call is restored immediately. The problem consists in coordinating choices such that either cell of the two beneficial ones is reached.

In more technical terms, coordination problems are characterized by at least two proper equilibria. If any of them is reached, no agent wishes to have acted differently. There is, in other words, no regret of choice. The

[34] See Lewis (1969); Ullmann-Margalit (1975); Brennan/Buchanan (1985); Menger (1963); Schotter (1981); Schelling (1960); Wärneryd (1990).

problem, hence, consists in coordinating choices to find an equilibrium in the absence of communication.

The above discussed examples pose in themselves a genuine problem of coordination, even if we neglect a possible time dimension[35]. Since one shot interactions, intrinsically interesting as they are, are of very limited importance for an understanding of social institutions in the world around us, let us consider recurrent choices and the role of norms and their emergence in reitinerated interaction.

Viewed technically, our actors could still judge each of the recurrent situations on its own merits and decide each time afresh, whether to call back or to wait, or in general terms, whether to stick to the pattern of behavior previously in operation or to depart from it. If coordination, however, has a history it becomes clear that previously adopted standards are of assistance in predicting the other party's choice and, consequently, in finding the right choice for oneself[36].

In recurrent situations, and this is an important point, conventions (rules) help to stabilize the expectations of the individuals involved. Many conventions and regularities in human behavior provide information, upon which our actors can rely in recurrent coordination problems, for example, how to dress for a specific occasion. More importantly for our purposes here, many institutions and rules, constitute in very much the same manner specific solutions to coordination problems. Examples of fundamental importance are language (being a device for coordination with regard to communication), money (being a universally accepted means of exchange), or some rules of the road[37].

Rules or conventions that solve coordination problems have specific properties well worth pondering. Coordination norms are self enforcing, since only concerted action will yield satisfactory payoffs. In turn, there is no gain attainable by unilateral defection. There is, at least in the pure case, no incentive to unilaterally deviate from standards that solve coordination problems. In case one player had acted differently, at least one actor would have been made worse off without improving the position of the other. It is neither beneficial to deviate from the established standard of how to behave in case a phone call is disconnected, nor to defect on a rule of grammar, nor to introduce one's own medium of exchange.

---

35 In a one shot interaction, choices are empirically quite often based on same salient properties of one equilibrium point, e.g., that it is a saddle point. See e.g., Schelling.

36 Brennan/Buchanan (1985) p. 8.

37 Whereas it is always true that these rules generate predictability as to the behavior of others, it is throughout possible that individuals seek advantage by cheating on rules that set reliable standards (say, fake money or improper scales and weighs). Here, the wrongdoer seeks to exploit the public trust on these standards.

2.1.2. Problems of Social Order as Coordination Problems

Let us pursue this line of reasoning by turning to specific problems of social order. Consider, as an example, the rules of the road. Each driver is confronted with the choice to adopt a rule of driving either to the right or to drive left.

A \ B	drive right	drive left
drive right	1, 1	−1, −1
drive left	−1, −1	1, 1

1.1

A \ B	right	left
right	2, 2	−1, −1
left	−1, −1	1, 1

1.2

A \ B	right	left
right	2, 1	−1, −1
left	−1, −1	1, 2

1.3

Matrix 1

In a first setting, we are going to analyze, we shall assume that either rule appears equally desirable to our players, as long as any rule is established (Matrix 1.1). In a second setting, let us suppose that the actors differ in their evaluation of possible rules that would solve the coordination problem, for example, that both players prefer right hand driving to left hand driving (Matrix 1.2). In the third matrix we depict choices such that, although coincidence of interests prevails, we do have some elements of conflict embodied in the coordination problem: Player A prefers right hand driving (say, because he is skilled in it), whereas player B comes from a country, where he was used to drive on the left. Cells I and IV of Matrix 1.3 illustrate the respective distributional advantages.

It is important to note that all the three matrixes illustrate genuine coordination problems. Given either of the diagonal cells, no player wants the other one to have acted differently. The interests of our players coincide primarily in installing a rule, any rule, that allows them to concert their choices and to reach mutually satisfactory outcomes. This property holds true, even though there might exist apparent elements of conflict, as in Matrix 1.3. Since coordination still remains superior to unilateral defection (the off-diagonals in terms of the matrix), individual defection, hence, is unattractive.

This last point draws our attention to a further property of rules that solve coordination problems. If a rule is established, there is a natural predilection towards its maintenance, even when some actors may strongly dislike its distributional impacts. Conformity to this rule is still the key to the pay-off of an equilibrium solution that is superior to any uncoordinated outcome. Moreover, this feature allows us to identify a certain potential for strategic interaction in the following sense. If the disfavored actor commits himself firmly to the alternative option (that would favor himself), even at the risk of incurring a short time utility loss, then he may induce his peer to switch to a different strategy[38] (namely the one that yields distributional advantages for himself).

Another point deserves our attention in this respect: The individuals need to share only the interest to establish a rule that solves the coordination problem they face. They need not share tasks, aims, and objectives they want to pursue under such rules. In fact, their objectives may be quite heterogenous. Consider again the case of driving: Individuals will hold a whole bundle of motives such as why to drive, when to drive and where to drive, and they may even be in conflict about the proper amount of individual driving. And yet, irrespective of their motives, they share an interest in the solution of the coordination game. Moreover, the structural analysis of coordination problems applies independently of the "moral" properties that are at stake. That means that it is equally likely or unlikely for individuals to coordinate their choices with regard to the decision to wait or to call back in cases of communication between the Mafia boss and a professional killer and two lovers.

[38] However, some distributional advantages that may accrue to one actor under a coordination norm may vanish if we allow for a time sequence. If individuals take all positions under a certain rule at random, benefits are equalized over time. As an example, consider the driving rule to give way to the car heading upwards. This rule poses a differential advantage on this driver, but since positions are switched as time moves on, there is in the long run no identifiable self interest to defend.

2.1.3. Conventions as Solutions to Recurrent Coordination Problems

We have focused so far on the functional role of rules in terms of solving recurrent coordination problems, but we have disregarded the production and emergence of such rules. Rules, such as right hand driving, can be designed by deliberate choice on the norm level, say, by agreement of the individuals involved (or statutory enactment). The constructivist approach to social institutions, however, is of limited explanatory power, as regards the actual emergence of rules, standards, and conventions that solve coordination problems. Some of the most fundamental social institutions, such as money and language, have emerged from a spontaneous evolutionary process. Acknowledging this fact alerts us to the need to assess the invisible hand explanations of social institutions[39].

What these explanations have in common is that they explain social institutions as the result of human action, but not of human design. Invisible hand approaches view rules as endogeneously emerging behavioral standards. Institutions and norms (conventions, rules) evolve as a systematic by-product of human action that is not deliberately meant to produce such norms. Spontaneous forces generate incentives that in themselves reconfirm tentatively emerging regularities in human behavior. Rules and norms, hence, are established without the process being consciously directed by any agent (or group of agents).

These regularities and standards of behavior emerge from interaction by anonymous individuals in large number environments. Yet, the process, by which these regularities gain foothold, does not require simultaneous action by all individuals that constitute the population. The emergence of a rule that solves the recurrent coordination problem is due to a snowballing effect[40]. Regularities of behavior start by the formation of clusters that after reinforcing themselves by endogenously generated incentives grow continuously. Once a critical mass is reached within the population, the behavioral standard or norm becomes stable.

There are several interesting features of spontaneously emerging conventions or rules that merit attention. First, it was shown in the literature that we can predict some rule, standard or convention to evolve in the absence of deliberately designed or installed rules. This is to say that we can expect either left hand driving or right hand driving to emerge, or in

[39] Such invisible hand approaches are associated most prominently with the work of Hayek and other scholars in the tradition of the Austrian School of Economics that experiences a remarkable revival in recent years. Furthermore, see Schotter (1981), Vanberg (1986), and Wärneryd (1990).

[40] See Sugden (1986) and Schotter (1981).

the example of telephone calls being interrupted either rule to evolve. We can, furthermore, predict a common language to emerge in recurrent communication problems. Such findings imply that we can, in general terms, predict coordination problems to be solved by spontaneous forces. Such rule would indeed emerge as a product of human action, but not of human design[41]. However, recognition of this fact, immediately draws upon a second finding: In this evolutionary process, there is an ambiguity involved in respect to the specific solution that will emerge. Whereas we can be sure that some rule that solves the coordination problem will evolve, we cannot determine its precise contents. Any rule that solves the coordination problem is feasible. In other words, nothing in the evolutionary process guarantees that the optimal rule (in the actors' own evaluations) will emerge[42]. Matrix 1 illustrates this property in terms of the rules of the road example. Whereas we can predict that either right hand driving or left hand driving (cell I or cell IV) will evolve spontaneously as a pattern of behavior, we will not be able to specify in advance the specific rule. Moreover, as regards a change in the factual circumstances, there are no inherent spontaneous forces that necessarily will yield an adaptation to the new environment. We may be stuck with the old, now inefficient solution.

In this respect, the notions of path dependencies, network externalities and evolutionary market failure are of considerable importance[43]. Perhaps the most prominent example is the maintenance of the QWERTY-keyboards for typewriters and computers. At the time of their development, these keyboards represented the leading technology. However, they remained dominant even at a time, when some superior keyboard arrangements had been developed. We cannot explain this phenomenon by relying (solely) on the technology effect that doubtlessly would have favored new keyboard-arrangements in terms of speed in writing. What has to be considered too is a network effect (or consumer externality) that represents the utility differences following from the number of consumers who adopt a certain technology (an important modern example being video-recorder systems). Insofar as these effects are taken into consideration, the development of technologies appears to be path-dependent. Technological solutions that are challenged by

41 The evolutionary finding of such solution may also, according to the relevant social problem, involve some severe loss (for instance in the example of rules on car driving).

42 The outcome will, inter alia, depend on the initial distribution of strategies in the population. If enough actors conform to a suboptimal rule, it will establish itself as the predominant rule and finally as a convention, without alternative rules being developed. However, a better convention also has a larger probability, other things being equal, that it will be converged upon from a random initial population; see e.g., Wärneryd (1990) pp. 100–102.

43 Blankart/Knieps (1993) and (1989); David (1985); Katz/Shapiro (1985) and (1986); Adams (1993).

superior alternatives and, hence, become suboptimal may resist the technological pressure because of their widespread diffusion. The evolutionary adoption of the superior technology may be foreclosed given the network externalities of the old solution. In the case of QWERTY-keyboards, typists were reluctant to invest in learning on the new keyboards in the absence of their sufficient spread, and producers could not successfully commence supplying the new technology in the absence of trained typists. In other words, the development was locked into QWERTY-arrangements, without superior solutions being able to establish themselves by spontaneous forces[44].

The basic message, as regards coordination problems, therefore is that they are inherently solvable: some, although not necessarily the optimal, solution to a coordination problem will be found by a spontaneous process. One caveat, however, applies in this respect. Whereas a solution to coordination problems will be found, if "snowballing" effects of gradually evolving standards are feasible, a different result emerges in the case of simultaneous coordination within recurrent n-person interaction. This problem has attracted only little attention so far[45].

Consider a game, in which each of the 1000 individuals involved, will receive 10$, if just 5 players separately write down the number 10 on a sheet of paper. If more than 5 players or less write down this number (to be determined ex ante), no one will receive anything. We may allow for repetitions of this game and for communication of the results achieved in the previous round (say, 24 individuals have written down the number 10). However, no communication must be allowed ex ante in order to coordinate strategies.

If individuals are identical and act simultaneously, I cannot see any inherent mechanism, by the use of which coordination could be achieved. Since outcomes are always ambiguous, no endogenously generated mechanism could steer outcomes in the right direction. Or in other words, no snowballing effects apply. If, for example, 100 players write down 10, no player knows whether he shall be the one to switch strategies for the next turn. Hence, coordination can only be achieved by chance. Consequently, the time required for reaching the equilibrium increases with the number of players involved.

There are few, if any examples of such n-person coordination games in real life settings[46]. One explanation, of course, could be the very fact that coordination is

44 Path-dependency in this understanding, however, does not in itself constitute a sufficient reason for the maintenance of inefficient solutions in each and every case. The technology effect can be so strong that it overwhelms by all means the existing network effect.

45 However, see Hirshleifer (1982).

46 Traffic jams, albeit incorporating PD-characteristics, come close to the type of

unlikely to be achieved and, therefore, is unlikely to be observed. If, however, communication is directly allowed for, or if individuals can even develop institutions that supplement for the lack of direct communication, such problem of coordination will be solved.

2.1.4. Implications: the Case for Institutional Design

There are three implications, worth considering, that stem from this analysis if we seek to assess the possibility for social order in the absence of some central agency. First, in the above mentioned n-person case, where snowballing effects are lacking, institutional design will be required to solve the coordination problem as such. Second, since the evolutionary emergence of a norm does not guarantee in itself optimality (to be noted again, in terms of the actors' evaluations), there remains a certain potential for improvement by deliberate action, that is by explicit choice on the norm level[47]. Third, such potential for constitutional improvement in a broad sense may occur, if some spontaneously emerged rules, which have previously been optimal, lose this property due to exogeneous developments or shocks. In such cases too, an evolutionary adaptation, a switch away from an established convention to a different type solution might be unlikely to occur[48]. Although a potential for deliberate constitutional reform to improve the performance of the rules in operation may exist and, moreover, is not unlikely to be expected on theoretical grounds, we have to be cautious with regard to policy. advice and policy action, even if the superiority of some new institutional arrangement may be evident. We have always to consider transitional costs that may prevent an improvement of the rules in operation, as in the case of left hand driving in England. Transitional costs, however, are small as regards a switch from imperial measures to the metric system in the US.

How to cure the maintenance of inefficient solutions? The considerations above concerning the QWERTY-keyboard arrangements suggest interventionist measures that dictate or at least coordinate (by announcing future standards) the move to the superior technology. However, once again, one has to be careful about the side effects of such measures. Since such interventions, if not conceived as open process

coordination problem, we consider here: Every morning commuters are caught in the same back-up. If the motorists could coordinate their trip to the town, such that some drive in earlier and others later, everybody could speed up. Yet, in real life settings, motorists are unable to communicate in any meaningful manner their choices. Hence, morning after morning coordination is not achieved.

47 Brennan/Buchanan (1985) p. 10.

48 See, in turn, the notion of constructive destruction in the work of Hayek and Schumpeter. These authors assert the tendency of superior solutions to dominate inferior standards by evolutionary pressure.

oriented regulation, eliminate the development of alternative solutions, optimal solutions could be excluded by possible premature interventionist regulations.

2.1.5. Coordination Problems and Self Enforcing Rules

Rules that solve coordination problems (irrespective of whether spontaneously generated or by deliberate choice) are, as already pointed out above, self reconfirming and self enforcing. With regard to rule adherence, their functional task coincides with the self interest of the individuals involved. Moreover, rules that solve coordination problems are largely self enforcing, even in situations that embody certain elements of conflict (as in Matrix 1.3). Such sources of conflict are overwhelmed by the mutual (or universal) interest in the coordinating rule[49].

Hence, in coordination problems we have a convergence of "individual" and "collective" interest, as to rule compliance. Moreover, we can illuminate the distinctive properties of coordination problems even better by adopting a constitutionalist perspective. In respect of coordination problems, an individual's constitutional interest (rule interest) and her compliance interests are in harmony. This assertion holds true, since the individual's choices on the action level are always conditional on the presumed behavior of her peers. Rules that solve coordination problems accomplish the functional task of providing information (and, hence, allowing legitimate expectations) that permits concerted action. Since unilateral defection does not pay, there are no incentives to depart from some behavioral standard, once adopted. In other words, agreement on a rule is sufficient to elicit compliance in the actual working of this rule.

We have arrived on an important conclusion. If all problems of social interaction were of coordination type, the romantic approach to society would be confirmed. It would suffice to establish certain rules or standards and individuals would conform with these standards, first because these rules accomplish their specific functional task, and, second, because it would individually be rational to comply. Unfortunately, this is not the case. Coordination problems are not the overall type of problems that we have to deal with when contemplating social interaction. There are other types of social interactions that follow a different structure, and, as explained below, in these settings elements of conflict prevail.

49 Occasional incentives for defection may exist though. Take the example of right hand driving. Here even the reliable driver may be prompted sometimes to break the rule. In pure cases of coordination problems (e.g., language), there are no incentives for deviation at all.

2.2. Prisoners' Dilemma Problems (2 by 2)

2.2.1. The Problem

We have, so far, considered settings of social interaction, whereby the actors' interests for coordinated choices coincide or at least are overwhelming. However, other choice settings exist, whereby the interests of the actors involved partly clash and partly coincide. This structure of social interaction is captured in the Prisoners' Dilemma analysis (henceforth: PD) which is a uniquely appropriate tool for discussing rational choices in situations[50] that capture the fundamental problem of social order.

Let me illustrate with examples. The first example is the original anecdote regarding the dilemma of two prisoners[51] who, accused of being accomplices to the same crime, are interrogated in separate cells. If both remain silent, evidence suffices only to convict them for some minor offense (say, unauthorized possession of firearms). If both plead guilty, both are convicted for the main charge, but receive, due to their collaboration with the Court, a reduced sentence (which still is worse than the minor offense). If, however, one prisoner confesses the crime and turns in his accomplice, he will go free (for having provided State's evidence), whereas his fellow receives a long term sentence. Given these constraints, namely absence of strategic communication and binding agreements, will either, neither, or both prisoners confess? If both confess, both get the reduced sentence for the main charge. However, both would be better off, if neither confessed (insofar the prisoners' interests coincide). Given that prisoner 2 remains silent, prisoner 1 could meliorate his position by confessing. So he will confess. Prisoner 2, considering now the option that his fellow could turn him in, will confess too, since he can, thereby, reduce his sentence. So both will end up in confessing, whereas both were better off, if they had remained silent.

In the real world, there are plenty of settings that fit the underlying structure of a PD. In particular, the problem of social order itself can be captured by means of a PD. Reduced to its barest essentials, the prisoners' dilemma illustrates in a "nutshell", why conflict prevails and social order is likely to collapse, notwithstanding that a potential for mutual improvement exists. Also, the arms race follows the logic of the PD. To avoid superiority of the adversary or to exploit his restraints, respectively, a party will defect on the disarmament agreement, previously agreed upon.

[50] J. Buchanan has made extensive use of this analytical tool in his studies on state and society, see his Limits of Liberty (1975) and pp. 64–68 in particular. For a general discussion see e.g., Luce/Raiffa (1967) pp. 88–113 and Rapoport (1966) pp. 123–144. Further Ullmann-Margalit (1977) pp. 18–73, Schotter (1981) and M. Taylor (1976).

[51] For the original story of the PD, see e.g., Luce/Raiffa (1967) pp. 94–95.

Due to the parallel motivations of both parties, they will end up, where they commenced.

A further illustration can be drawn from two competing firms in an oligopoly market setting. If either of the firms considers cutting its price and underselling the competitor, it would be advantageous for firm 1 to undersell firm 2, and it would be disastrous to be undersold. So firm 1 will cut its prices. However, since this is also the reasoning of firm 2, it will resort to the same strategy. Both end up with reduced prices, though it would have been advantageous for both to remain at the old price level.

Bearing in mind the above examples, it becomes clear, upon reflection, that, whereas all the examples concern problems of mutual "cooperation", they differ sharply in their moral dimension. Whereas the PD captures the dependencies of choices that constitute the dilemma, it does not answer in itself the normative issue, whether cooperation in terms of the PD is valued as a "good". In the arms race case, for example, we seek to secure the cooperative solution that is essentially synonymous with peaceful coexistence.

In the original PD story and the oligopoly market example, the cooperative solution is, of course, beneficial for the actors, but unwanted from a "moral perspective": We want prisoners to confess and firms to compete. In these cases, the dilemma that plagues the players involved, is valued positively by the rest of the community (the law abiding community and the consumers, respectively). Whether the cooperative solution in the PD is "good" or "bad" depends, hence, on the group that we consider and, consequently, on a normative standard of evaluation that we have to evoke from outside. It follows, moreover, that any theory of constrained utility maximization in terms of the PD framework is in itself not equivalent to a theory of moral behavior.

Bearing that in mind, we can derive a clear cut policy implication. Insofar as PDs of the first type are concerned, the normative goal is to hinder the players from reaching the cooperative outcome. One can accomplish such goal either by withholding enforceability or by directly prohibiting "cooperative" behavior. Conversely, as regards the second type of PD, our interest lies in establishing and encouraging cooperation. Let us now turn to the paradigmatic case of such PD, the problem of social order.

2.2.2. Prisoners' Dilemma and Social Order

2.2.2.1. The Hobbesian Problem in a "Nutshell"

Let me now discuss in some detail the choices of individual actors in context of the Hobbesian problem of social order[52]. Let us assume two players, A

[52] However, we still remain in the realm of a single encounter, a one shot

and B, in the state of nature, where they do not respect each other's property rights. Each player considers two options (to cooperate or to defect)[53]. Both players prefer a situation of mutual cooperation to mutual defection. Matrix 2 illustrates the choices that our players face and gives the utility payoffs they can obtain.

		B: cooperate	B: defect
A	cooperate	1, 1	−2, 2
	defect	2, −2	0, 0

Matrix 2

In this game, actor A is confronted with the following choice: If B cooperates, A can yield a higher payoff by defecting. Unilateral defection offers the most advantageous position attainable, since A could perfectly exploit B. However, A will also consider the option that B will defect on his part. Given this perspective, A would run the risk of exploitation if he himself cooperated. The best strategy A could resort to, if B defects, is to defect on his part. On the overall, A will always be better off by defecting, no matter what B does. In other words, defection is unconditionally preferred by A. It embodies his dominant strategy.

Moreover, if we recast this setting in terms of desirability of cells: The best outcome for A is a vector of strategies, when B cooperates and A himself defects. Second ranks mutual cooperation. Third best is mutual defection, and the worst outcome constitutes unilateral cooperation, whereby cooperator A is exploited by defecting B. Hence, defection is A's optimal choice. However, the same kind of reasoning is of course true for B as well. Regardless of what A does (whether he defects or cooperates), B will always prefer to defect. It is this unfortunate property, namely dominance of defection, that constitutes the genuine dilemma. Thus, both

[53] We may interpret the features of this setting as those of a Hobbesian Warre (or of two superpowers involved in strategic choices within an arms' race). Given this situation, each player can either refrain from his prior belligerent policy (he can cooperate in a broad sense), say, by commencing a peaceful peasant life, or he can pursue his aggressions and violate B's property (he can defect in broad sense). Thus, player A and B can either choose a cooperation or a defection strategy. If both cooperate, both enjoy a relatively peaceful living (and save costs for defense). If both defect, i.e., pursue their aggressive strategies (construct new weapons etc.), they remain in the status quo.

players will end up in the mutual defection cell, although both players prefer mutual cooperation over mutual defection[54]; it is the only cell that is an equilibrium, an outcome, where there is no regret of choices.

Being a Nash-equilibrium, the mutual defection cell is the only stable outcome. Given the strategy chosen by his partner, no player has an incentive to switch to a different strategy. The reasons are: By a unilateral switch to cooperation, our actor will be exploited. Universal defection, hence, constitutes the only equilibrium[55]. Whereas the mutual defection cell constitutes the only equilibrium, it is the only cell that is not pareto-optimal. Both players could be made better off by a move to cell one. Though a move to the cooperation/cooperation cell is pareto-optimal, this pareto-optimal position is inaccessible by individual rational choices.

2.2.2.2. Extensions: Benevolent Players

One could object, however, that the analysis suggested above hinges crucially on the "unmoral" features of the players, namely the actors' unconstrained self interest. Such critique asserts that, if we abandoned the assumption that players always try to exploit each other, we would get a different result. This, however, is not the case.

Let us, thus, reassess the one-shot PD by assuming A to be a nice player who is not interested in harming and exploiting B[56]. A, however, is not willing to suffer exploitation by B. Will A cooperate? Not necessarily. A cannot single out cells (outcomes); he can only choose rows. The only choice for A that does not harm B is to cooperate. However, in case A cooperates, he exposes himself to exploitation by B. Given this perspective, A will be likely to adopt a maximin strategy: He will try to choose a strategy such that if the worst outcome occurs, his losses are minimized. However,

[54] Dominance of defection is the salient point in PD interaction. In the original PD story, prisoner A is always better by confessing. If B remains silent, A can turn him in and go free. If B confesses, A can avoid a full sentence by doing likewise. Similarly, price cutting strictly dominates maintenance of the old price schedule.

[55] All other cells do not constitute equilibria. Both in the case of mutual cooperation (Cell I) and unilateral defection (Cell II and Cell III), the other party can always improve its performance by resorting to defection. Consequently, such moves will ultimately generate Cell IV as outcome (which is mutual defection). From this defection/defection cell mutual cooperation is inaccessible by individual behavior. The actors are locked into the defection/defection cell by their individual choices.

[56] For this argument, let us assume the payoffs of the original PD unaltered. If we, however, change payoffs in the sense that the temptation to exploit the cooperating opponent is eliminated, there is no single dominant strategy and we have two equilibrium points. This game, assurance, is not discussed here, see, e.g., Sen (1974) p. 59 and Ullmann-Margalit (1977) p. 35.

only defection guarantees the highest minimum payoff. Hence, he must, for the sake of avoiding exploitation, resort to defection. In other words, defection is A's defense strategy against potential wrongdoers. Even if A is a nice actor, he will likely defect on his turn, since only defection grants him insurance against the highest possible loss.

More generally, we can divide defectors into two groups that differ substantially in motivation but not at all in the kind of strategy, they finally resort to. There are actors who primarily defect to take advantage of their co-player, that is to exploit their cooperative peers. Defectors of this kind are primarily attracted by the enticing extra payoff of unilateral defection as given in cell II (type 1 defection). The second group of defectors may be described as "benevolent" defectors. They do not care about exploiting their partner, they even might be highly uneasy about the possibility that they take advantage of the other party involved in case of unilateral defection. However, they are reluctant to expose themselves to exploitation. They defect to protect themselves against the worst possible outcome (type 2 defection).

Furthermore, let us suppose that both individuals involved in the PD encounter are nice players who would like to cooperate provided only that there is no risk of exploitation. However, as long as they remain ignorant as to their peers' decisive characteristics, they will still, depending on their degree of risk-aversion and the likelihood of their partner being a defector, resort to defection as their defense strategy[57].

Moreover, A will still defect against B, if he knows B to be a cooperator, but (for what reason ever) assumes that B mistakenly takes himself (A) for a defector. This analysis implies that our benevolent actors may not cooperate, as long as there is lack of credible signals for cooperation. It follows that even benevolent persons are unlikely to produce any better results than unconstrained utility maximizers[58].

[57] See in this respect A. Sen's assertion that the PD will be inherently solvable, if each player behaved, as if he had preferences as in the assurance game and had the assurance of similar behavior by his partner; Sen (1974) p. 60.

[58] Since in this last case considered here (both players are cooperators in search for a better assurance than defection), mutual defection is due to asymmetric information, we will get a different result, if we release the assumption of simultaneous action. Should it be possible that the players involved act successively, the one who commences can precommit himself to cooperation, signaling that he is a cooperator. Harvesting aid may serve as an example: Let us assume that B's fields are ready for harvest a week before A's: In case A helped B this week, B could turn down A's request for support next week and thus perfectly exploit A. Given this perspective, A will be reluctant to help B. However, if A supposes B to cooperate, given that A is a cooperator, he will simply assist B in his harvesting efforts. Thus in the set of cases now assessed mutual coöperation can emerge even in a one-shot interaction. Conversely, as long as asymmetric information is not overcome, two cooperators are locked into a defection/defection cell.

2.2.3. Recurrent PDs

So far, we have confined our analysis to one shot interaction. This focus of attention on the most elementary conceptual PD interaction, the two-by-two one shot case, finds it justification in the fact that such simple framework already captures most of the fundamental structural problems of social order. Still, individuals act in time and it is this time dimension that requires institutional solutions to (recurrent) PD interaction[59].

Provided for infinite repetitions, cooperation may spontaneously emerge in PD settings under certain conditions. However, infinite repetitions are a rare event and, hence, finite repetitions are more rewarding for further analysis. Unfortunately, even a finite repetition of the game does not alter the fact that defection is the dominant strategy. If we have in the PD supergame a certain number of replications, we can expect rational individuals to resort to defection in the last encounter, since there is no future to account for. The choice setting is wholly analogous to the one shot case. Backward induction, however, prompts defection also in the second last encounter. However, if this is true, the same type of reasoning applies in each other replication of the game. Hence, defection will be the overall strategy in the game.

However, under certain conditions cooperations may evolve though. As we shall see, cooperation may gain foothold in recurrent PD interaction on the basis of reciprocity[60] or, more generally, of conditional cooperation. We will turn to this question in section 3 below.

2.2.4. Rules Interests and Compliance Interests: Prisoners' Dilemma Interaction and (Non-Enforced) Contracts

Having identified, first, the defection/defection cell as being not pareto-optimal and, second, mutual cooperation as the pareto-superior outcome, we can conclude that there must be a way for improvement, in terms of a mutually agreed move, to cell I. Recognition of this structure is central for

[59] Since much of what I shall discuss in the section below on enforcement will revolve around recurrent PDs, some sketchy remarks shall suffice here.

[60] We can relate this issue to the above discussed question of the spontaneous emergence of norms. Since individual rational choice dictates defection as an individual's dominant strategy, it is unlikely if we remain strictly within the model that spontaneous forces generate rules that solve PD problems. Whereas spontaneous forces can bring about rules and conventions that solve coordination problems, an evolutionary emergence of norms is less likely in PD settings (Vanberg 1986). To be more specific, there is a certain potential for selective incentives being produced in recurrent PD interaction, where reciprocity features are involved. Since, much of the analysis in the section on enforcement will be devoted to the reciprocity issue, I shall postpone considerations on this subject to this point.

the whole constitutionalist contractarian approach. It reveals that there is a possibility for agreement on institutional improvement to establish mutual cooperation as the outcome of social interaction.

However, as we shall see, PDs have the unfortunate property that their logic does not change, even if we allow the players to engage in a non-enforced contractual arrangement. Assume that our players acknowledge the structure of the game they are in. They, therefore, fully recognize that the move to the mutual cooperation cell would be beneficial for both. Let us suppose, further, that they, hence, are in perfect agreement on a disarmament contract that is designed to trace a concerted way out of the self chosen state of mutual defection. However, even if the players agreed on a cooperative solution, the system will collapse as before. A (non-enforced) disarmament contract only reflects the players' intent to reach a stage of mutual cooperation, but it will not, in itself, induce them to cooperate. The incentives on the action level have not altered[61]. In other words, each player's compliance interests remain the same. Defection is still each player's dominant strategy. If player A supposes B will honor the agreement, then he can exploit him by defecting. If B should violate the agreement on his part, the best A can do is to resort to defection likewise. The reciprocal incentive structure again applies for B. Mutual defection will plunge the system back into its original position. Driven by their own self interest, both players will defect on the very rule on which they have agreed (out of the same self interest) previously.

Recognition of this fact reveals a point of general importance. We have to distinguish carefully between an individual's interests on the norm level, and his interests for compliance. Insofar as PD-like interaction is concerned, the individual's "constitutional interests" and his action interests are not in harmony. Norms that are designed to solve PD settings do not in themselves provide incentives for compliance. Since defection remains each individual's dominant strategy, the whole system will result into a state of mutual defection.

The negative account regarding voluntary cooperation in PD settings does not hinge on the claim that an individual's public interests (some inclination towards the "public good") are countered by some "personal" self interest in the strict sense. There is no need to distinguish between different motivations of the same individual. All what is required here, is to acknowledge the conflict of the same individual's interests on different stages of choice. The individual agrees on the contract, driven by his own personal interest, and he defects on this contract, driven by the same self

[61] See for a discussion of the constitutional versus the action level Vanberg/Buchanan (1988) and more generally Brennan/Buchanan (1985).

interest. PDs illustrate the systematic gap between the interest of the individuals in rule making (on the constitutional level) and on their action/compliance-level. This property of PD settings implies that aims of overcoming the inherent instability of mutually beneficial cooperation in such situations are essentially not self stabilizing and self- enforcing.

Given these properties of PD settings, we can now reconsider the structural differences between coordination problems and PD problems. Coordination problems and PD problems have in common the fact that the players involved prefer a state of mutual cooperation over a state of mutual defection. However, for PD interaction, overall cooperation is inherently unstable. Whereas in coordination problems mutual cooperation constitutes an equilibrium, the only equilibrium in PD interaction is a state of overall defection[62].

Insofar as coordination problems are concerned, the interests of all actors coincide, or at least their interests for coordination are overwhelming. All players are faced with conditional choices, depending on what they expect the other(s) to do. PD settings, in turn, are characterized by unconditional choices; no matter what the others do, it is always better to defect.

Rules (conventions, standards) that solve recurrent coordination problems, hence, are self enforcing, and they are self enforcing, because they are equilibria. The players' norm interests and their action interests coincide. If a coordination equilibrium is reached, no agent wishes any other agent to have acted differently. Pure coordination rules, hence, lack any incentive for unilateral defection. Rules that solve PD settings, in turn, seek to shape individual choices towards uniformity to escape the only (unsatisfactory) equilibrium. Since the contents of such rules aim at establishing a non-equilibrium position as outcome, such rules cannot be self enforcing. Each individual can meliorate his position by violating the standards agreed upon. The only equilibrium still is mutual defection. In other words, whereas the individuals' constitutional and action interests coincide in respect to coordination norms, they are in conflict as regards PD interaction. Constitutional choices in PD interaction are always jeopardized by the incentives on the action level for defection. In other words, enforcement plays an important role in PD interaction and a negligible role in coordination problems.

Since there is the potential for institutional improvement, individuals should be able to devise means of eliminating the gap between rules interests and compliance interests. It is this fundamental insight that captures in a nutshell the core of the contractarian approach to social order. Most importantly, rational individuals can employ coercive devices and enforced contracts to ensure and stabilize the cooperative outcome.

[62] Both in coordination problems and PD interaction, if an equilibrium is reached, no player wishes to have acted himself differently. Yet, in coordination games no player in the equilibrium wishes any other player to have acted differently, whereas in PD settings each player wishes the other to cooperate since he himself would then yield the higher payoff of unilateral defection.

One feasible device is, hence, the use of a set of sanctions that back the rules agreed upon. In case of deviance, these sanctions are to be imposed on the rule-breaker. Penalties can dramatically change the payoffs of the players, providing by their preventive (deterrent) effects the crucial incentives for compliance on the action level. The threat of being sanctioned when breaking the law leads to an internalization of those negative external effects that the wrongdoer would otherwise exert. Therefore, individuals may agree in their own self interest on the establishment of an agent to enforce the law by coercion. The implementation of such a contract is beneficial to everybody, even though some of the consenting individuals will be punished under those rules, they had agreed upon previously.

Acknowledgment of this possibility for agreement does not imply that individuals will seek the protection by the state regardless of the costs involved. A successful accomplishment of the enforcement task needs funding, and hence, requires some scheme of tax payments that furnish the resources needed for this enterprise. Under reasonable assumptions, however, the costs saved by the erection of a monopoly of power (the modern state) exceed those required for maintaining some kind of "criminal justice system".

Skogh/Stuart (1982) have developed a simple model in which they try to outline the conditions under which agreement regarding such costly enforcement (and the installment of an enforcing agency) will occur. The social contract considered in this model involves 4 components, namely, a rule establishing property rights, sanctions for violations, the erection of a criminal justice system, and a rule embodying the tax schedule. Vis a vis an Hobbesian equilibrium, the social contract may secure overall gains in terms of reduced expenditures for defense and predation, but imposes costs in terms of tax payments. As long as this social enterprise is, by and large, economically beneficial, it is rational to engage into an enforced social contract. Such contract, hence, is on the overall agreeable.

2.2.5. The Incentive Gap and Self Management

2.2.5.1. General Observations

One way of solving the central problem of social interaction is to rely on an enforcing agent who, empowered with a monopoly of legitimate threat and force ensures rule adherence by coercion. Individuals, obviously, do not establish such coercive entity because a monopoly of power is attractive as such, but because they lack more enticing alternatives.

Alternative institutional regimes may be attractive simply because of costs. The funding of the institutional apparatus of the state involves costs that could be (at least partially) avoided, if there were some feasible alternative that can accomplish the same task without the need of a huge state apparatus. However, this need not be the only reason. Since the state

by its very nature embodies coercion, some individuals experience an intrinsic disutility when living under such enforcing institutions.

Taylor has addressed this question in terms of the original PD setting[63]. His argument runs along the following lines: Individuals prefer Cell I (mutual cooperation) vis a vis Cell IV (overall defection). Individuals, hence, agree on institutional arrangements that bring about Cell I as an outcome. However, this analysis does not tell us anything of how individuals evaluate the role of the state as the enforcer of the cooperative outcome. It might well be that the mere existence of this enforcer in itself changes payoffs and reduces the utility that individuals may be able to enjoy. Individuals may derive a higher utility in a setting of universal cooperation, in which it is not the state that enforces the rules agreed upon.

Are there any feasible alternatives that could accomplish the task of providing security, but do not involve those disadvantages that the existence of a state necessarily implies[64]? Given the structural gap between constitutional and compliance interests in PD interaction, we have to analyze the potential of the individual to bridge this gap by means of self management in a broad sense. I will address this issue, which is of particular relevance with respect to the feasibility of cooperation in a world of type 2 defectors, in terms of the "weakness of will problem".

2.2.5.2. Self Management and Precommitments

A sole reliance on willpower may be too weak a device to overcome the incentive gap involved in any PD interaction. However, the individual in her long term interests may rationally develop a mechanism to put certain options (say, defection in a broad sense) beyond her own reach on the action level, because she knows that she will, at the moment of actual choice, fall into temptation and prefer current utility to her long term interests. In other words, individuals may rationally adopt measures to cope with their own myopia.

Such myopia exists because people act in time and discount the future. The higher their discount rate, the more attached they are to the present. If we model an individual such that she embodies time preferences (for the present), then she will prefer the current situation's short term benefits over higher long term returns. If an individual discounts the future in a hyperbolic (more than exponential way), she will counter her long term (constitutional) interests by her own choices on the

---

63 Taylor (1976).

64 In some societies formal rules and sanctions are largely substituted by social sanctions (e.g., Nigeria), in others a coherent system of religious norms, beliefs and creeds constitute the protection of individual rights.